# A constructive proof of the phase-type characterization theorem

Illés Horváth[1] and Miklós Telek[2,3]

[1] MTA-BME Information Systems Research Group, Budapest, Hungary

[2] Department of Telecommunications, Technical University of Budapest, Budapest, Hungary

[3] Inter-University Center of Telecommunications and Informatics, Debrecen, Hungary

e-mail: horvath.illes.antal@gmail.com, telek@hit.bme.hu

February 3, 2015

**Abstract**

The paper presents a new proof of O'Cinneide's characterization theorem [7]. It is much simpler than the original one and constructive in the sense that we not only show the existence of a phase type representation, but present a procedure which creates a phase type representation. We prove that the procedure succeeds when the conditions of the characterization theorem hold.

Keywords: Matrix-Exponential distribution, phase-type distribution, vector-matrix representation, randomization.

## 1 Introduction

The characterization theorem of O'Cinneide [7] proves that any finite order matrix exponential function which is strictly positive in $(0, \infty)$ and satisfies the dominant eigenvalue condition has a finite dimensional phase type (PH) representation. Based on this theorem Mocanu and Commault [2] proposed a procedure for computing the PH representation of such matrix exponential function. A quite different approach from Maier [3] proposes a similar procedure based on Soittola's automata-theoretic algorithms [8]. All of these papers prove the characterization theorem, but use complex mathematical concepts, such as polytopes, or positive rational sequences. Additionally, both procedures in [2] and in [3] are implicit in the sense that an essential parameter ($\tau$ in [2] and $c$ in [3]) are found as a result of a numerical search.

In this paper we present a constructive proof of the characterization theorem by proposing an explicit procedure for computing a phase type (PH) representation of a matrix exponential function and showing that the procedure always terminates successfully if the matrix exponential function satisfies the positivity and the eigenvalue conditions.

Compared to the existing resuls, one of the main advantages of the presented constructive proof is that it is rather elementary, using basic function and matrix theory and stochastic interpretation of Markov processes. Another contribution of the paper is that it links the sparse monocyclic representation [2] to the characterization theorem [7].

## 2 Preliminaries

**Definition 1.** *Let $X$ be a non-negative random variable with probability density function (pdf)*

$$f_X(x) = \frac{\mathrm{d}}{\mathrm{d}x}\Pr(X < x) = -\boldsymbol{\alpha}\mathbf{A}e^{\mathbf{A}x}\mathbb{1}, \quad x \geq 0,$$

*where $\boldsymbol{\alpha}$ is an initial row vector of size $n$ with $\int_0^\infty f_X(x)dx = 1$ (there is no probability mass at zero), $\mathbf{A}$ is a square matrix of size $n \times n$ and $\mathbb{1}$ is the column vector of ones of size $n$. In this case, we say that $X$ is matrix-exponentially distributed with representation $\boldsymbol{\alpha}, \mathbf{A}$, or shortly, ME($\boldsymbol{\alpha}, \mathbf{A}$)-distributed.*

In Definition 1 the elements of $\boldsymbol{\alpha}$ and $\mathbf{A}$ are real numbers without any specific restriction on their sign and the only restriction is that $f_X(x)$ is non-negative for $x \geq 0$. We note that $\mathbf{A}$ and $e^{\mathbf{A}x}$ commute.

For a given representation $(\boldsymbol{\alpha}, \mathbf{A})$, the size $n$ of $\boldsymbol{\alpha}$ (and $\mathbf{A}$) is called the *order* of the representation. The representation of a given ME distribution is not unique.

**Theorem 1.** *[1] Let ME($\boldsymbol{\alpha}, \mathbf{A}$) of order $n$ and ME($\boldsymbol{\gamma}, \mathbf{G}$) of order $m$ be two ME distributions with pdf $f_X(x)$ and $f_Y(x)$, respectively. If $\mathbf{A}$ is $n \times n$*

*If there exists a matrix $\mathbf{W}$ of cardinality $n \times m$ such that*

$$\boldsymbol{\gamma} = \boldsymbol{\alpha}\mathbf{W}, \quad \mathbf{AW} = \mathbf{WG}, \quad \mathbb{1}_n = \mathbf{W}\mathbb{1}_m,$$

*then ME($\boldsymbol{\alpha}, \mathbf{A}$) $\equiv$ ME($\boldsymbol{\gamma}, \mathbf{G}$) (that is, $f_X(x) = f_Y(x)$).*

*Proof.*

$$f_Y(x) = -\boldsymbol{\gamma}\mathbf{G}e^{\mathbf{G}x}\mathbb{1}_m = -\boldsymbol{\alpha}\mathbf{W}\mathbf{G}e^{\mathbf{G}x}\mathbb{1}_m = -\boldsymbol{\alpha}\mathbf{A}e^{\mathbf{A}x}\mathbf{W}\mathbb{1}_m = -\boldsymbol{\alpha}\mathbf{A}e^{\mathbf{A}x}\mathbb{1}_n = f_X(x).$$

$\square$

Theorem 1 will be used as a representation transformation tool. The size of column vector $\mathbb{1}$ is explicitly indicated in the theorem as a subscript.

**Definition 2.** *A representation of an ME distribution has minimal order if the distribution has no representation of a smaller order. A representation of minimal order is referred to as a* minimal representation.

In a minimal representation, there are no "extra" or "redundant" eigenvalues in matrix $\mathbf{A}$. More precisely a minimal representation has the following properties [10]:

P1) All Jordan blocks of $\mathbf{A}$ have different eigenvalues.

P2) All eigenvalues contribute to the distribution with maximal multiplicity. For example, a Jordan block of size $n_i$ corresponding to eigenvalue $-\lambda_i$ results in the terms $\sum_{j=1}^{n_i} c_{\lambda_i,j} x^{j-1} e^{-\lambda_i x}$ in $f_X(x)$, where $c_{\lambda_i,n_i} \neq 0$.

P3) $\boldsymbol{\alpha}$ is not orthogonal to any of the right-eigenvectors of $\mathbf{A}$.

P4) $\mathbb{1}$ is not orthogonal to any of the left-eigenvectors of $\mathbf{A}$.

P5) The Jordan block structures of all minimal representations of an ME distribution are identical.

These properties are explained further in Appendix B. Based on these properties, a minimal representation can be obtained directly from $f_X$. If $f$ takes the form

$$f(x) = \sum_{i=1}^{m} \sum_{j=1}^{n_i} c_{\lambda_i,j} x^{j-1} e^{-\lambda_i x}$$

where $\lambda_i$ are different and $c_{\lambda_i,n_i} \neq 0$, then we will consider the following representation $(\boldsymbol{\alpha}, \mathbf{A})$:

$$\mathbf{A} = \begin{pmatrix} \mathbf{J}_1 & 0 & \dots & & 0 \\ 0 & \mathbf{J}_2 & 0 & \dots & 0 \\ \vdots & & & & \\ 0 & \dots & & 0 & \mathbf{J}_m \end{pmatrix},$$

where

$$\mathbf{J}_i = \begin{pmatrix} \lambda_i & 1 & \dots & & 0 \\ 0 & \lambda_i & 1 & \dots & 0 \\ \vdots & & & & \\ 0 & \dots & & 0 & \lambda_i \end{pmatrix}.$$

and $\mathbf{J}_i$ is of size $n_i$. $\boldsymbol{\alpha}$ can be calculated by solving

$$-\boldsymbol{\alpha}e^{\mathbf{A}x}\mathbf{A}\mathbb{1} = f_X(x);$$

this equation can be solved because the left-hand side contains all the terms $x^{j-1}e^{-\lambda_i x}$ up to $j \leq n_i$ for $i = 1, \dots, m$.

**Lemma 2.** *The representation* $(\boldsymbol{\alpha}, \mathbf{A})$ *is minimal for* $f_X$.

The proof is essentially due to properties P1-P5 and the fact that no Jordan block of size smaller than $n_i$ can represent the term $x^{n_i-1}e^{-\lambda_i x}$. Appendix B elaborates more on this topic.

If representation $(\boldsymbol{\alpha}, \mathbf{A})$ is minimal then there are some straightforward necessary conditions for vector $\boldsymbol{\alpha}$ and matrix $\mathbf{A}$ to define a valid distribution:

C1) The eigenvalues of $\mathbf{A}$ have negative real part (to avoid divergence of $f_X(x)$ as $x \to \infty$).

C2) There is a real eigenvalue of $\mathbf{A}$ with maximal real part (to avoid oscillations to negative values as $x \to \infty$).

C3) $\boldsymbol{\alpha}\mathbb{1} = 1$ (normalizing condition which ensures $\int_0^\infty f_X(x)dx = 1$).

C4) If for all $i \in \{0, 1, \ldots, j-1\}$ the $i$th derivative of $f_X(x)$ is zero then the $j$th derivative of $f_X(x)$ is non-negative (to avoid decreasing behavior around $x = 0$).

If any of these necessary conditions are violated then the tuple consisting of the vector $\boldsymbol{\alpha}$ and matrix $\mathbf{A}$ does not define a valid ME distribution. Note that non-minimal representations might contain any additional eigenvalues, including for example positive ones.

A subclass of ME distributions is the class of phase-type distributions (PH distributions).

**Definition 3.** *If $X$ is an ME($\boldsymbol{\alpha}, \mathbf{A}$) distributed random variable, where $\boldsymbol{\alpha}$ and $\mathbf{A}$ have the following properties:*

- $\alpha_i \geq 0$, $\boldsymbol{\alpha}\mathbb{1} = 1$

- $A_{ii} < 0$, $A_{ij} \geq 0$ for $i \neq j$, $\mathbf{A}\mathbb{1} \leq 0$

- $\mathbf{A}$ *is non-singular,*

*then we say that $X$ is phase-type distributed with representation $(\boldsymbol{\alpha}, \mathbf{A})$, or shortly, PH($\boldsymbol{\alpha}, \mathbf{A}$) distributed.*

PH distributions can be interpreted as the time of absorption in a CTMC [6] and consequently the conditions of Definition 3 are sufficient for vector $\boldsymbol{\alpha}$ and matrix $\mathbf{A}$ to define a valid distribution. Vector $\boldsymbol{\alpha}$ or matrix $\mathbf{A}$ satisfying the conditions of Definition 3 are referred to as *Markovian*.

The following properties are essential for the characterization of ME distributions.

**Definition 4.** *An ME($\boldsymbol{\alpha}, \mathbf{A}$) distribution satisfies the* dominant eigenvalue condition *(DEC) if for some minimal representation ME($\boldsymbol{\gamma}, \mathbf{G}$), $\mathbf{G}$ has a single eigenvalue with maximal real part. This eigenvalue is called the dominant eigenvalue. Its multiplicity may be higher than 1.*

Definition 4 excludes the case when $a$ is the dominant real eigenvalue and there is a pair of complex eigenvalues with the same real part, for example $a \pm \mathbb{i}b$, where $\mathbb{i}$ is the imaginary unit.

Properties P1-P5 ensure that if C1-C4 hold for one minimal representation ME($\boldsymbol{\alpha}, \mathbf{A}$), they hold for all equivalent minimal representations. Additionally, if the dominant eigenvalue has multiplicity higher than 1, then it belongs to a Jordan-block whose size is equal to the multiplicity of the dominant eigenvalue.

**Definition 5.** *The ME($\boldsymbol{\alpha}, \mathbf{A}$) distribution with density $f_X$ satisfies the* positive density condition *if $f_X(x) > 0$ for all $x \in (0, \infty)$.*

**Theorem 3.** *[7] If $f_X$ is ME($\boldsymbol{\alpha}, \mathbf{A}$) distributed, then $f_X$ has a finite dimensional PH($\boldsymbol{\beta}, \mathbf{B}$) representation iff the following two conditions hold:*

- *ME($\boldsymbol{\alpha}, \mathbf{A}$) satisfies the dominant eigenvalue condition;*

- *$f_X$ satisfies the positive density condition.*

The original proof of O'Cinneide in [7] is rather complex, using Laplace-Stieltjes transform and geometric properties of the space of PH-distributions. In this paper we present an algorithm that gives a constructive and altogether more elementary proof, using function and matrix theory.

# 3  Procedure and proof

Our main goal is an algorithm that provides a constructive proof for the sufficient direction of Theorem 3, that is, given that the dominant eigenvalue condition and the positive density condition hold for $\mathrm{ME}(\boldsymbol{\alpha}, \mathbf{A})$, find a PH-representation equivalent to $\mathrm{ME}(\boldsymbol{\alpha}, \mathbf{A})$; in other words, find a vector-matrix pair $(\boldsymbol{\beta}, \mathbf{B})$ where $\boldsymbol{\beta}$ and $\mathbf{B}$ are Markovian and define the same distribution as $\mathrm{ME}(\boldsymbol{\alpha}, \mathbf{A})$.

This section is devoted to the algorithmic construction, also stating the theorems used along the way. Proofs are given in Appendix C.

We also included a proof for the necessary direction of Theorem 3 in Appendix B. While the proof of the necessary direction is straightforward using the techniques in [7], we opted to include a self-contained, elementary proof that is more in line with the methods of the present paper.

## 3.1  Sketch of the algorithm

The algorithm consists of five main steps. Steps 1 and 2 are preparatory, and Step 5 is just correction related to Step 2.

- Step 1. We find an equivalent minimal representation $(\boldsymbol{\alpha}_1, \mathbf{A}_1)$ for $(\boldsymbol{\alpha}, \mathbf{A})$ if it is not minimal by eliminating any "extra" eigenvalues of $\mathbf{A}$, which does not contribute to the pdf. We refer to Lemma 2 and [1] for a straightforward and computationally stable method of finding a minimal representation.

- Step 2. This step applies only if density is zero at 0, that is, $f_X(0) = 0$. This step is essentially what may be called "deconvolution": we represent $f_X$ as the convolution of some $f_Y$ matrix exponential density function with $f_Y(0) > 0$ and an appropriate Erlang-distribution $\mathrm{Erlang}(k, \mu)$ (see Lemma 4); if $f_Y$ has a Markovian representation, then it gives a straightforward Markovian representation for $f_X$ as well (see Lemma 5). Thus we only need to find a Markovian representation for $f_Y$ (and the corresponding representation, which is obtained from Lemma 4), where $f_Y(0) > 0$. If this step is applied, Steps 3 and 4 are applied for $f_Y$ instead of $f_X$, and we switch back to $f_X$ in Step 5.

- Step 3. An equivalent representation $(\boldsymbol{\gamma}, \mathbf{G})$ is given with Markovian matrix $\mathbf{G}$, while $\boldsymbol{\gamma}$ may still have negative elements. The main tool of this step is the so-called monocyclic structure (with Feedback-Erlang blocks). Typically, the size of $\mathbf{G}$ is larger than that of $\mathbf{A}_2$ (because each pair of complex conjugate eigenvalues is represented with at least 3 phases); that said, $\mathbf{G}$ is a sparse matrix with a simple block bi-diagonal structure. For this step only the dominant eigenvalue condition is necessary.

- Step 4. $\boldsymbol{\gamma}$ and $\mathbf{G}$ are transformed further into $\boldsymbol{\beta}$ and $\mathbf{B}$ where $\boldsymbol{\beta}$ is Markovian (and the Markovity of $\mathbf{B}$ is also preserved) essentially by adding an "Erlang-tail" (a number of sequentially connected exponential phases with identical rates) of proper order and rate to the monocyclic structure described by the Markovian matrix $\mathbf{G}$. The main mathematical tool of this step is the approximation of elementary functions. Essentially, this last step is the contribution of the paper. The skeleton of this step is composed of the following elements:

  - Find $\tau$ such that $\boldsymbol{\gamma} e^{\mathbf{G}\tau} > 0$ (element-wise). Such $\tau$ always exists if the dominant eigenvalue and the positive density conditions hold and the pair $(\boldsymbol{\gamma}, \mathbf{G})$ results from the previous step. We remark that for a general representation, even if $\mathbf{G}$ is Markovian, such a $\tau$ may not exist. This is further explained after Lemma 9.

  - Find $\lambda'$ such that
  $$\boldsymbol{\gamma} \left( \mathbf{I} + \frac{\mathbf{G}}{\lambda} \right)^{\tau\lambda} > 0 \quad \forall \lambda \geq \lambda'$$
  which is always possible since $\left\| \boldsymbol{\gamma}(\mathbf{I} + \frac{\mathbf{G}}{\lambda})^{\tau\lambda} - \boldsymbol{\gamma} e^{\mathbf{G}\tau} \right\| \to 0$ as $\lambda \to \infty$.

  - Let $\epsilon = \inf_{x \in (0, \tau)} f_X(x)$. $\epsilon > 0$ because of the positive density condition and the result of Step 2. Find $\lambda''$ such that
  $$\left| -\boldsymbol{\gamma} e^{\mathbf{G}\tau} \mathbf{G} \mathbb{1} + \boldsymbol{\gamma} \left( \mathbf{I} + \frac{\mathbf{G}}{\lambda} \right)^{\tau\lambda} \mathbf{G} \mathbb{1} \right| < \epsilon \quad \forall \lambda \geq \lambda''.$$
  This ensures that $-\boldsymbol{\gamma} \left( \mathbf{I} + \frac{\mathbf{G}}{\lambda} \right)^{k} \mathbf{G} \mathbb{1} > 0$ for $k = 1, \ldots, n$ where $n = \tau\lambda''$. This is always possible when $\epsilon > 0$.

4

– Extend the $(\boldsymbol{\gamma}, \mathbf{G})$ representation with an Erlang tail of rate $\lambda = \max(\lambda', \lambda'')$ and order $n = \lambda\tau$.

- Step 5. If Step 2 was applied, at this point we have a Markovian representation for $f_Y$. To switch back to $f_X$, we use Lemma 5. If Step 2 was not applied, Step 5 does not apply either.

## 3.2   Step 1: Minimal representation

Starting from representation $(\boldsymbol{\alpha}, \mathbf{A})$, we can obtain a minimal representation $(\boldsymbol{\alpha}_1, \mathbf{A}_1)$ with the application of a representation minimization method. A minimal representation can be obtained through several approaches. One possibility is directly from the pdf $f(x) = \boldsymbol{\alpha}e^{\mathbf{A}x}\mathbb{1}$ as in Lemma 2. Another, computationally stable order reduction method is the Staircase method from [1], which uses singular value decomposition. In any case, the minimal representation $(\boldsymbol{\alpha}_1, \mathbf{A}_1)$ enjoys properties P1-P5.

There are two important properties that can be determined from a minimal representation (or the density function directly). These are the value and the multiplicity of the dominant eigenvalue and the validity of the dominant eigenvalue condition. We denote the dominant eigenvalue (which is real and negative) by $-\lambda_1$ and its multiplicity by $n_1$. Indeed, $\lambda_1$ and $n_1$ determine the asymptotic rate of decay of the pdf: it decays like $c_{\lambda_1,n_1}x^{n_1-1}e^{-\lambda_1 x}$, where $c_{\lambda_1,n_1}$ is a positive constant, more precisely

$$\lim_{x\to\infty} \frac{f(x)}{x^{n_1-1}e^{-\lambda_1 x}} = c_{\lambda_1,n_1}.$$

## 3.3   Step 2: Positive density at zero

In the case when $\mathrm{ME}(\boldsymbol{\alpha}_1, \mathbf{A}_1)$ is such that $f_X(x) > 0$ for positive $x$ values, but $f_X(0) = 0$, then based on the following lemma, we represent $\mathrm{ME}(\boldsymbol{\alpha}_1, \mathbf{A}_1)$ as the convolution of an Erlang distribution and a matrix exponential distribution $\mathrm{ME}(\boldsymbol{\alpha}_2, \mathbf{A}_2)$ whose density is positive at 0. Actually, it turns out from the proof of the following lemma that $\mathbf{A}_1 = \mathbf{A}_2$.

**Lemma 4.** *If $f_X(x) = -\boldsymbol{\alpha}_1 e^{\mathbf{A}_1 x}\mathbf{A}_1\mathbb{1}$ is a matrix exponential pdf with*

$$f_X(x) > 0 \quad \forall x > 0, \qquad f_X^{(i)}(x)\Big|_{x=0} = 0 \quad i = 0,\dots,l-1, \qquad f_X^{(l)}(x)\Big|_{x=0} > 0, \qquad (1)$$

*then $f_X$ can be written in the form*

$$f_X = f_Y * g(l, \mu, \cdot),$$

*for some large enough $\mu$, where $g(l, \mu, x) = \frac{\mu^l x^{l-1}e^{-\mu x}}{(l-1)!}$ is the Erlang$(l, \mu)$ pdf, $*$ denotes convolution and $f_Y(x)$ is a matrix exponential function with*

$$f_Y(x) > 0 \quad \forall x \ge 0.$$

The proof of Lemma 4 is given in the Appendix. The representation $(\boldsymbol{\alpha}_2, \mathbf{A}_2)$ can be constructed either from $f_Y$ via Lemma 2 or by using the fact that $\mathbf{A}_1 = \mathbf{A}_2$ and calculating $\boldsymbol{\alpha}_2$ from the appropriate linear equations.

Lemma 4 and the following composition ensures that $f_X(x)$ and $f_Y(x)$ have a Markovian representation and satisfy the dominant eigenvalue condition at the same time if $\mu > \lambda_1$.

**Lemma 5.** *If $f_Y(x)$ is ME distributed with representation $(\boldsymbol{\alpha}_2, \mathbf{A}_2)$ of order $m$ and $\mu > \lambda_1$ then*

$$f_X(x) = f_Y(x) * g(l, \mu, x),$$

*is ME distributed with initial vector $\boldsymbol{\beta} = \{1, 0, 0, \dots, 0\}$ and generator matrix*

$$\mathbf{B} = \begin{pmatrix} -\mu & \mu & & \\ & \ddots & \ddots & \\ & & -\mu & \mu\boldsymbol{\alpha}_2 \\ & & & \mathbf{A}_2 \end{pmatrix},$$

*where the first $l$ blocks of the matrix are of size one and the last block is of size $m$. Additionally, if $(\boldsymbol{\alpha}_2, \mathbf{A}_2)$ is Markovian then $(\boldsymbol{\beta}, \mathbf{B})$ is Markovian as well.*
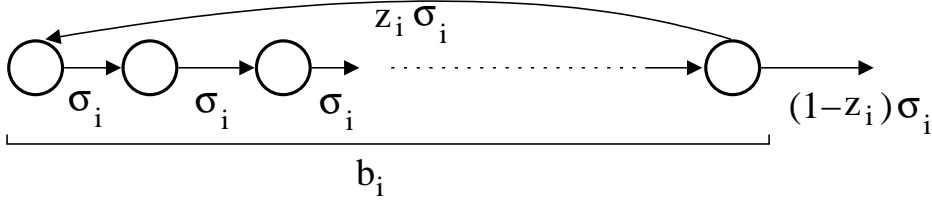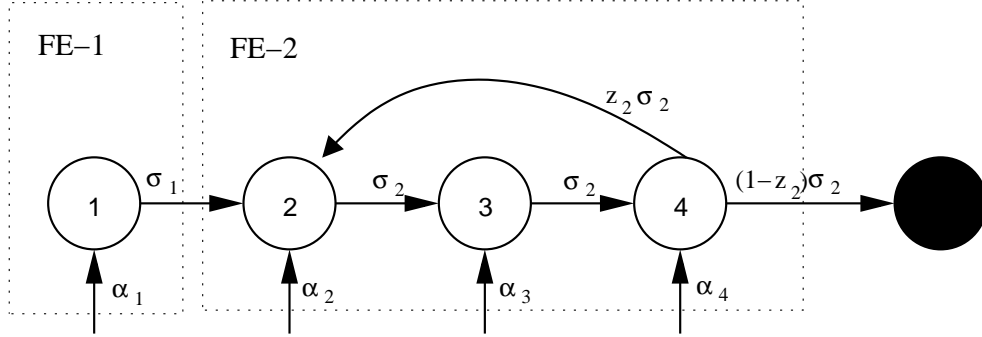
Figure 1: FE-diagonal block.



Figure 2: FE-diagonal representation of a generator with a real eigenvalue ($\sigma_1$) and a pair of complex ones.

*Proof.* Based on the structure of $\mathbf{B}$, the time to leave the first $l$ phases is Erlang($l, \mu$) distributed and the time spent in the set of phases from $l+1$ to $m$ is ME($\boldsymbol{\alpha}, \mathbf{A}$) distributed. □

Based on Lemma 4 and 5 it remains to prove that the matrix exponential density function $f(x)$ with $f(0) > 0$ satisfying the dominant eigenvalue and the positive density conditions has a Markovian representation.

## 3.4   Step 3: Markovian generator

The aim of this subsection is to transform the potentially non-Markovian representation $(\boldsymbol{\alpha}, \mathbf{A})$ of a ME distribution to a representation $(\boldsymbol{\gamma}, \mathbf{G})$ where $\mathbf{G}$ is a Markovian transient generator matrix satisfying the properties of the matrix of a PH distribution (Definition 3). For matrix $\mathbf{G}$, we apply the matrix structure proposed in [5]. It is a block bi-diagonal matrix structure, where each block represents a real eigenvalue or a pair of complex conjugate eigenvalues of $\mathbf{A}$. The blocks associated with real eigenvalue $-\lambda_i$ ($-\lambda_i < 0$) are of size one, the diagonal element is $-\lambda_i$ and the first sub-diagonal element is $\lambda_i$. The blocks associated with complex eigenvalues are composed by Feedback-Erlang (FE) blocks.

**Definition 6.** *[5] A* Feedback-Erlang (FE) block *with parameters* $(b, \sigma, z)$ *is a chain of $b$ states with transition rate $\sigma$ and one transition from the $b$th state to the first state, with rate $z\sigma$ (c.f. Figure 1). The probability $z \in [0, 1)$ is called the feedback probability.*

A FE block $(b, \sigma, z)$ with length $b = 1$ and $z = 0$ corresponds to a real eigenvalue $-\sigma$ and is referred to as *degenerate* FE blocks. Matrix $\mathbf{G}$ contains as many FE blocks (degenerate or non-degenerate) associated with a real eigenvalue or a pair of complex conjugate eigenvalues as the multiplicity of the eigenvalue. A non-degenerate FE block where $b$ is odd has a real eigenvalue and $(b-1)/2$ complex conjugate eigenvalue pairs. A non-degenerate FE block where $b$ is even has 2 real eigenvalues and $(b-2)/2$ complex conjugate eigenvalue pairs. In both cases the eigenvalues are equidistantly located on a circle in the complex plane around $-\sigma$. The dominant eigenvalue of the FE block (the one with the largest real part) with parameters $(b, \sigma, z)$ is always real and equals to $r = -\sigma\left(1 - z^{\frac{1}{b}}\right)$ [5]. Denote the eigenvalues of matrix $\mathbf{A}$ by $-\lambda_j$; the dominant eigenvalue (which is real) is $-\lambda_1$. The FE blocks representing the eigenvalues are composed as follows

6

- if $\lambda_j$ is real, the corresponding FE block is a degenerate block; thus the parameters are:

$$\sigma_j = \lambda_j, \quad b_j = 1, \quad z_j = 0,$$

- if $\lambda_j = a_j \pm \mathrm{i}c_j$ $(a_j > \lambda_1 > 0, c_j > 0)$ is a complex conjugate pair, the parameters are:

$$b_j = \left\lceil \frac{2\pi}{\left| \pi - 2\arctan\left(\frac{c_j}{-\lambda_1 + a_j}\right) \right|} \right\rceil,$$

$$\sigma_j = \frac{1}{2}\left( -2a_j - c_j \tan\frac{\pi}{b_j} + c_j \cot\frac{\pi}{b_j} \right),$$

$$z_j = \left( 1 - \left( -a_j - c_j \tan\frac{\pi}{b_j} \right)/(2\sigma_j) \right)^{b_j},$$

where $\lceil x \rceil$ denotes the smallest integer greater than or equal to $x$.

This construction of the FE blocks ensures that $\lambda_1$ remains the dominant eigenvalue of matrix $\mathbf{G}$, that is, the dominant eigenvalue of any FE block $(r_j)$ is less than $-\lambda_1$ except the one(s) associated with $-\lambda_1$.

Connecting the obtained FE blocks such that the exit transition of an FE block (whose rate is $\lambda_j(1 - z_j)$, see Figure 1, in case of non-degenerate FE block and $\lambda_j$ in case of a degenerate one) is connected to the first state of the next FE block composes a block bi-diagonal matrix (c.f. Figure 2). The obtained matrix $\mathbf{G}$ is Markovian and its Jordan form contains all Jordan blocks of matrix $\mathbf{A}$. We order the FE blocks such that the first $n_1$ FE blocks are the $n_1$ degenerate FE blocks associated with $-\lambda_1$. The order of the rest of the FE blocks are irrelevant. The FE blocks based finite Markovian representation of the eigenvalues of $\mathbf{A}$ is always feasible when the dominant eigenvalue condition holds. If there was a pair of complex conjugate eigenvalues $a_j \pm \mathrm{i}c_j$ which violates the dominant eigenvalue condition such that $a_j = \lambda_1$ then the denominator of $b_j$ would be zero.

Figure 2 depicts an example of a Markovian generator which is the monocyclic representation of a generator with a dominant real eigenvalue $(-\lambda_1 = -\sigma_1)$ and a pair of complex conjugate eigenvalues in FE-diagonal form. In this representation there are two FE blocks, one of length $b_1 = 1$ with rate $\sigma_1$, and one of length $b_2 = 3$ with rate $\sigma_2$ and feedback probability $z_2$. The associated generator matrix is

$$\mathbf{G} = \left( \begin{array}{c|ccc} -\sigma_1 & \sigma_1 & 0 & 0 \\ \hline 0 & -\sigma_2 & \sigma_2 & 0 \\ 0 & 0 & -\sigma_2 & \sigma_2 \\ 0 & z\sigma_2 & 0 & -\sigma_2 \end{array} \right).$$

In order to find an equivalent representation of $\mathrm{ME}(\boldsymbol{\alpha}, \mathbf{A})$ with matrix $\mathbf{G}$ we need to compute vector $\boldsymbol{\gamma}$, for which $\mathrm{ME}(\boldsymbol{\alpha}, \mathbf{A}) \equiv \mathrm{ME}(\boldsymbol{\gamma}, \mathbf{G})$, with the help of Theorem 1. Let $n$ and $m$ $(n \leq m)$ be the order of $\mathbf{A}$ and $\mathbf{G}$, respectively. Compute matrix $\widehat{\mathbf{W}}$ of size $n \times m$ as the unique solution of

$$\mathbf{A}\widehat{\mathbf{W}} = \widehat{\mathbf{W}}\mathbf{G}, \qquad \widehat{\mathbf{W}}\mathbb{1} = \mathbb{1},$$

and based on $\widehat{\mathbf{W}}$ vector $\boldsymbol{\gamma}$ is

$$\boldsymbol{\gamma} = \boldsymbol{\alpha}\widehat{\mathbf{W}}.$$

Since $\mathbf{G}$ is Markovian, the obtained $(\boldsymbol{\gamma}, \mathbf{G})$ representation is Markovian if $\boldsymbol{\gamma}$ is non-negative, but this is not necessarily the case. The case when $\boldsymbol{\gamma}$ has negative elements is considered in the following subsection.

## 3.5 Step 4: Markovian vector

At this point in the algorithm, the ME distribution is described by representation $(\boldsymbol{\gamma}, \mathbf{G})$ of order $u$ which has a block bi-diagonal, Markovian matrix $\mathbf{G}$, and a vector $\boldsymbol{\gamma}$ with at least one negative element. In the next

step we extend the $(\boldsymbol{\gamma}, \mathbf{G})$ representation with an additional $n$ phases in the following way.

$$
\mathbf{B} = \begin{pmatrix}
\mathbf{G} & -\mathbf{G}\mathbb{1} & & & \\
& -\lambda & \lambda & & \\
& & & \ddots & \ddots \\
& & & & -\lambda
\end{pmatrix}, \tag{2}
$$

where $\mathbf{B}$ is of order $u + n$ (the size of the upper left block of $\mathbf{B}$ is $u$, the remaining $n$ blocks are of size one). $-\mathbf{G}\mathbb{1}$ is a non-negative column vector of size $u$. Due to the structural properties of $\mathbf{G}$ it contains exactly one non-zero element, which is the last element and it contains the exit rate from the last FE block. The transformation matrix $\mathbf{W}$ of size $u \times (u + n)$, which transform from representation $(\boldsymbol{\gamma}, \mathbf{G})$ to representation $(\boldsymbol{\beta}, \mathbf{B})$ is the unique solution of $\mathbf{GW} = \mathbf{WB}$, $\mathbf{W}\mathbb{1}_{u+n} = \mathbb{1}_n$. Fortunately, due to the special structure of matrix $\mathbf{B}$, $\mathbf{W}$ is rather regular.

**Lemma 6.** $\mathbf{W}$ *has the following form:*

$$
\mathbf{W} = \left( \ \left(\mathbf{I} + \tfrac{\mathbf{G}}{\lambda}\right)^n \ \middle| \ \left(\mathbf{I} + \tfrac{\mathbf{G}}{\lambda}\right)^{n-1} \tfrac{-\mathbf{G}\mathbb{1}}{\lambda} \ \middle| \ \left(\mathbf{I} + \tfrac{\mathbf{G}}{\lambda}\right)^{n-2} \tfrac{-\mathbf{G}\mathbb{1}}{\lambda} \ \middle| \ \ldots \ \middle| \ \tfrac{-\mathbf{G}\mathbb{1}}{\lambda} \ \right),
$$

*where the size of the first block is $u \times u$, the size of the remaining blocks is $1 \times u$.*

*Proof.* Substituting this expression of $\mathbf{W}$ into $\mathbf{GW} = \mathbf{WB}$ and $\mathbf{W}\mathbb{1}_{u+n} = \mathbb{1}_n$ results in identities. $\qquad \square$

Our goal is to find $n$ and $\lambda$ such that $\boldsymbol{\beta} = \boldsymbol{\gamma}\mathbf{W}$ is Markovian (that is non-negative), where

$$
\boldsymbol{\gamma}\mathbf{W} = \left( \ \boldsymbol{\gamma}\left(\mathbf{I} + \tfrac{\mathbf{G}}{\lambda}\right)^n \ \middle| \ \boldsymbol{\gamma}\left(\mathbf{I} + \tfrac{\mathbf{G}}{\lambda}\right)^{n-1} \tfrac{-\mathbf{G}\mathbb{1}}{\lambda} \ \middle| \ \boldsymbol{\gamma}\left(\mathbf{I} + \tfrac{\mathbf{G}}{\lambda}\right)^{n-2} \tfrac{-\mathbf{G}\mathbb{1}}{\lambda} \ \middle| \ \ldots \ \middle| \ \boldsymbol{\gamma}\tfrac{-\mathbf{G}\mathbb{1}}{\lambda} \ \right). \tag{3}
$$

The first block of this vector is of size $u$ and the remaining $n$ blocks are of size 1. We need to prove that this vector is nonnegative for an appropriate pair $(\lambda, n)$.

**Theorem 7.** *There exists a pair $(\lambda, n)$ such that $\boldsymbol{\gamma}\mathbf{W}$ is strictly positive.*

The rest of this subsection is devoted to proving Theorem 7. We assume everything that was done so far, for example that the dominant eigenvalue condition and the positive density condition hold, the density is positive at zero and also that the matrix $\mathbf{G}$ is Markovian and in monocyclic form such that the degenerate FE block(s) representing the dominant eigenvalue $-\lambda_1$ are the first one(s). First we present a heuristic argument, then the formal proof.

### 3.5.1 Heuristic argument

$\lambda$ and $n$ are typically chosen to be large (see [5]). However, finding an appropriate pair is not as simple as choosing some large $\lambda$ and a large $n$. For each $n$, the set of appropriate values of $\lambda$ forms a finite interval. If $n$ is large enough, this interval is nonempty, but – without further considerations – it is impossible to identify this interval (or even one element of it). Vice versa, for each $\lambda$ there is a finite set of appropriate values for $n$. This means that the naive algorithm of increasing the values of $n$ and $\lambda$ – without further considerations – may possibly never yield an appropriate pair. For this reason, we instead propose a different parametrization, which takes the dependence between $n$ and $\lambda$ into account better.

Let $\tau = n/\lambda$. $\tau$ turns out to be a value interesting in its own right. The ME pdf resulting from the pair $(\boldsymbol{\gamma}\mathbf{W}, \mathbf{B})$ has a term coming from the first block of $\mathbf{B}$ and it has $n$ terms coming from the Erlang-tail. We argue that the terms coming from the Erlang-tail can be regarded as an approximation of the original pdf on the interval $[0, \tau]$, while the term coming from the first block is some sort of correction that makes the approximation exactly equal to the original pdf. Each of the terms in the Erlang-tail contribute an Erlang pdf with rate $\lambda$ and order $k \in [1, \ldots, n]$ to the pdf. The Erlang$(\lambda, k)$ pdf is concentrated around the point $\frac{k}{\lambda} = \frac{k\tau}{n}$. These points are situated along the interval $[0, \tau]$ in an equidistant way with distance $\frac{1}{\lambda}$.

The weight (initial probability) of the Erlang pdf centered around the point $\frac{k}{n}\tau$ is

$$
\boldsymbol{\gamma}\left(\mathbf{I} + \frac{\mathbf{G}}{\lambda}\right)^k \frac{-\mathbf{G}\mathbb{1}}{\lambda} \approx \boldsymbol{\gamma} e^{\frac{k\tau}{n}\mathbf{G}} \frac{-\mathbf{G}\mathbb{1}}{\lambda} = \frac{1}{\lambda} f_X\left(\frac{k\tau}{n}\right),
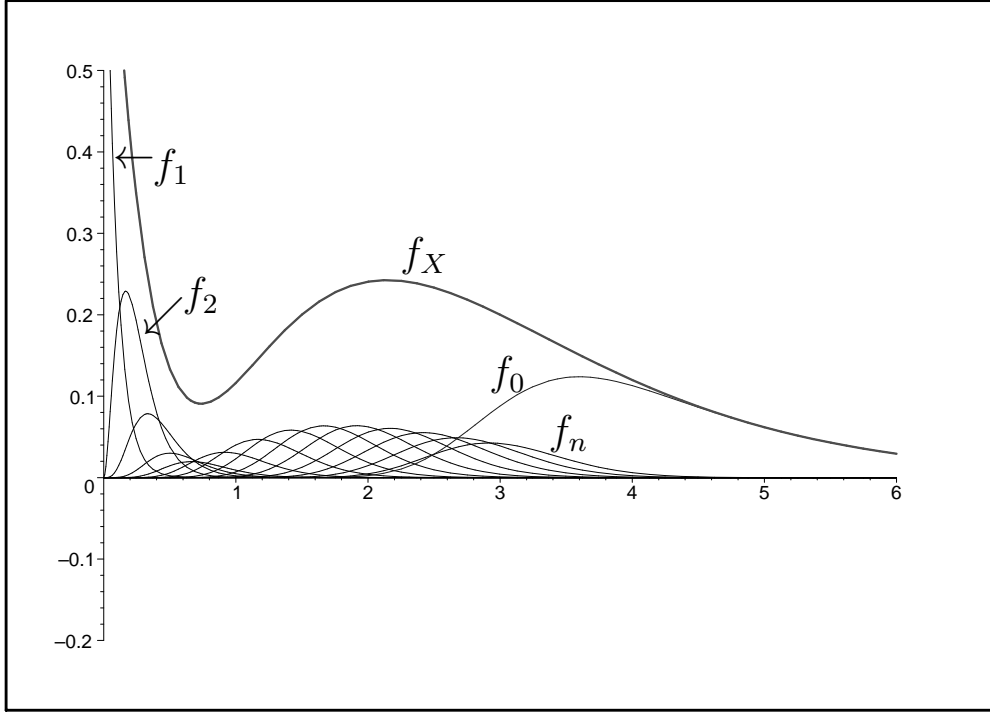$$

Figure 3: Erlang pdf's approximating the original one

which means that the weights are approximately equal to samples of the original pdf at points $\frac{k\tau}{n}$, $k \in [1, \ldots, n]$ divided by $\lambda$, resulting in a pdf that is approximately equal to the original along the interval $[0, \tau]$.

The first block of $\boldsymbol{\gamma}\mathbf{W}$ is different. From the form of $\mathbf{B}$ it is clear that the contribution of the first block is concentrated after the point $\tau$; the role of this block is essentially to make a correction in the interval $[\tau, \infty]$, where the previous Erlang-approximation does not hold.

Altogether the previous argument can be depicted nicely in Figures 3 and 4. We denote

$$f_k(x) = \boldsymbol{\gamma}\left(\mathbf{I} + \frac{\mathbf{G}}{\lambda}\right)^k \frac{-\mathbf{G}\mathbb{1}}{\lambda} \; g(k, \lambda, x), \quad k = 0, \ldots, n-1$$

the approximating Erlang terms and

$$f_0(x) = \boldsymbol{\gamma}\left(\mathbf{I} + \frac{\mathbf{G}}{\lambda}\right)^n e^{-\mathbf{G}x}(-\mathbf{G}\mathbb{1}) * g(n, \lambda, x)$$

the correction term. In Figure 3, the approximating Erlang terms roughly follow the graph of $f_X$, while $f_0$ is concentrated after $\tau$. (The values are $\tau = 3, \lambda = 12$ and $n = 36$; to make the figure visually apprehensible, only some of the approximating Erlang functions were included with slightly increased weights.)

The value of $\lambda$ controls how concentrated the approximating Erlang pdf's are and also controls how close their weights are to the sampling of the original pdf. Given that $f_X(x) > 0$ for $x > 0$, this means that for *any* choice of $\tau$, the Erlang-approximation has positive weights if $\lambda$ is large enough. The choice of $\tau$ is only important to make sure that the weights assigned to the correction term are also positive. Figure 4 shows an example where $\lambda$ is too small (notably $\lambda = 4$). In this case, some of the approximating Erlang functions have negative coefficients.

### 3.5.2 Formal proof

Before the actual proof, some results are stated as standalone lemmas. Their proofs are in Appendix C.

The first one is essentially a real approximation, so we state it in that form too, along with the matrix version which is useful for our purposes. Relevant properties of matrix (and vector) norms can be found in Appendix A.
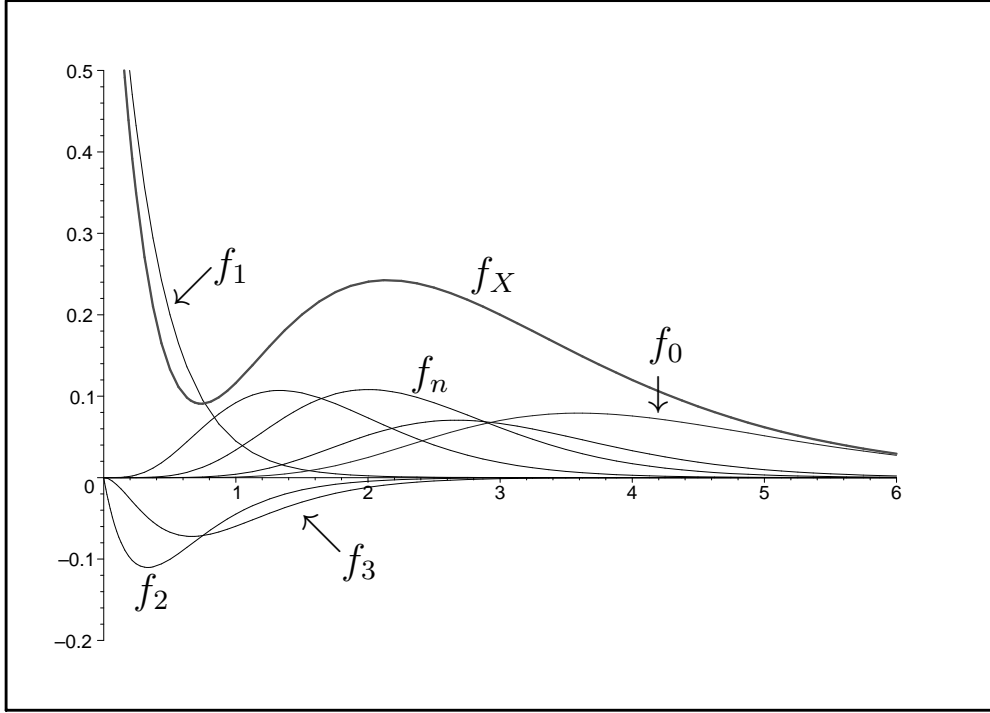
Figure 4: If $\lambda$ is too small, some Erlang pdf's are negative

**Lemma 8.** *i) For any fixed $r > 0$ and positive integer $n$,*

$$\sup_{|z| \leq r} \left| e^z - \left(1 + \frac{z}{n}\right)^n \right| \leq \frac{r^2 e^r}{2n},$$

*and the supremum is obtained at $z = r$.*

*ii) For any $\mathbf{H}$ square matrix,*

$$\left\| e^{\mathbf{H}} - \left(\mathbf{I} + \frac{\mathbf{H}}{n}\right)^n \right\| \leq \frac{r^2 e^r}{2n},$$

*where $r = \|\mathbf{H}\|$.*

We state one more lemma. It identifies the main terms in $e^{\mathbf{G}x}$ when $\mathbf{G}$ is in monocyclic form.

**Lemma 9.**

$$
\begin{aligned}
\left(e^{\mathbf{G}x}\right)_{1j} &\sim C_j x^{j-1} e^{-\lambda_1 x} \quad && \textit{if } 1 \leq j \leq n_1 \\
\left(e^{\mathbf{G}x}\right)_{1j} &\sim C_j x^{n_1-1} e^{-\lambda_1 x} \quad && \textit{if } n_1 < j \leq u, \\
\lim_{x \to \infty} \frac{\left(e^{\mathbf{G}x}\right)_{ij}}{\left(e^{\mathbf{G}x}\right)_{1j}} &= 0 \quad && \textit{if } \quad 2 \leq i \leq u, \quad 1 \leq j \leq u,
\end{aligned}
$$

*where $C_j$ denote positive (combinatorial) constants and $f(x) \sim g(x)$ denotes that $\lim_{x \to \infty} f(x)/g(x) = 1$. The last relation means that the first row dominates all other rows as $t$ tends to infinity.*

Note that the last part of Lemma 9 is stated as $\frac{\left(e^{\mathbf{G}x}\right)_{ij}}{\left(e^{\mathbf{G}x}\right)_{1j}} \to 0$; in fact, the elements $\left(e^{\mathbf{G}x}\right)_{ij}$ are in a form similar to $\left(e^{\mathbf{G}x}\right)_{1j}$, just with either the same exponential term and lower degree polynomial terms, or lower exponent (and in this case, the polynomial term does not matter). The actual exponents and polynomial terms, along with the constants $C_j$ can be calculated explicitly from the proof of Lemma 9, but will not be used.

10

We emphasize that Lemma 9 relies heavily on the monocyclic structure of $\mathbf{G}$, notably on the fact that the upper bi-diagonal elements (elements $(1,2), (2,3), \dots$) of the matrix are strictly positive.

Now we are ready to prove Theorem 7.

**Proof of Theorem 7.**

We assume that the matrix exponential density function $f_X$ associated with representation $(\boldsymbol{\gamma}, \mathbf{G})$ satisfies $f_X(0) > 0$, the dominant eigenvalue and the positive density conditions, and that $\mathbf{G}$ is in monocyclic block structure with the first block corresponding to the dominant eigenvalue $\lambda_1$.

*First we show that the first coordinate of $\boldsymbol{\gamma}$, denoted by $\gamma_1$, is positive.*

If $\gamma_1 = 0$, then the multiplicity of $-\lambda_1$ is $n_1 - 1$ according to the structure of matrix $\mathbf{G}$ (see (7) in the proof of Lemma 9 in subsection C.3), which is in conflict with the fact that the multiplicity of $-\lambda_1$ in the minimal representation is $n_1$.

$f_X(x)$ is dominated by the first row of $e^{\mathbf{G}x}$ for large values of $x$ and consequently the sign of $f_X(x)$ is determined by $\gamma_1$. The elements of $e^{\mathbf{G}x}$ are transient probabilities of the Markov chain with generator $\mathbf{G}$, consequently they are non-negative. The elements of the first row of $e^{\mathbf{G}x}$ are strictly positive for $x > 0$ because the FE-blocks are connected that way that all states are reachable from the first state (cf. Figure 2). According to Lemma 9 $f_X(x)$ is dominated by the first row of $e^{\mathbf{G}x}$ for large values of $t$ and consequently the sign of $f_X(x)$ is determined by $\gamma_1$. More precisely, Lemma 9 implies that

$$0 < f_X(x) = \boldsymbol{\gamma}(-\mathbf{G})e^{\mathbf{G}x}\mathbb{1} \sim C\lambda_1\gamma_1 x^{n_1-1}e^{-\lambda_1 x}$$

where $C = \sum_{j \geq n_1} C_j > 0$ and $\lambda_1 > 0$.

*Next we show that there exists a $\tau$ such that $\boldsymbol{\gamma}e^{\mathbf{G}\tau}$ is positive.*

For the first row of $\boldsymbol{\gamma}e^{\mathbf{G}x}$ we have

$$\left(\boldsymbol{\gamma}e^{\mathbf{G}x}\right)_{1j} \sim C_j\gamma_1 x^{j-1}e^{-\lambda_1 x} \qquad\qquad \text{if } j < n_1,$$

$$\left(\boldsymbol{\gamma}e^{\mathbf{G}x}\right)_{1j} \sim C_j\gamma_1 x^{n_1-1}e^{-\lambda_1 x} \qquad\qquad \text{if } n_1 \leq j \leq u,$$

from Lemma 9. Thus $\boldsymbol{\gamma}e^{\mathbf{G}x}$ is positive if $x$ is large enough. For a constructive procedure to find $\tau$, one can double $x$ starting from $n_1/\lambda_1$ as long as $\min(\boldsymbol{\gamma}e^{\mathbf{G}x}) < 0$. It is not necessary to find the smallest $x$ for which $\boldsymbol{\gamma}e^{\mathbf{G}x}$ is nonnegative.

*After that we show that there exists $\lambda'$ such that $\boldsymbol{\gamma}(\mathbf{I} + \frac{\mathbf{G}}{\lambda})^{\lambda\tau} > 0$ for $\lambda \geq \lambda'$.*

Apply Lemma 8 with $\mathbf{H} = \mathbf{G}\tau$ and $n = \lambda\tau$ to get that

$$\left\|\left(\mathbf{I} + \frac{\mathbf{G}}{\lambda}\right)^{\lambda\tau} - e^{\mathbf{G}\tau}\right\| \to 0$$

as $\lambda \to \infty$, and consequently

$$\left\|\boldsymbol{\gamma}\left(\mathbf{I} + \frac{\mathbf{G}}{\lambda}\right)^{\lambda\tau} - \boldsymbol{\gamma}e^{\mathbf{G}\tau}\right\| \to 0,$$

meaning that $\boldsymbol{\gamma}(\mathbf{I} + \frac{\mathbf{G}}{\lambda})^{\lambda\tau}$ is also strictly positive if $\lambda$ is large enough. Let $\epsilon_1 = \min(\boldsymbol{\gamma}e^{\mathbf{G}\tau})$; in accordance with Lemma 8, define $\lambda'$ as the solution of

$$\|\boldsymbol{\gamma}\|\frac{(g\tau)^2 e^{g\tau}}{2\lambda\tau} = \epsilon_1. \tag{4}$$

where $g = \|\mathbf{G}\|$. Then $\boldsymbol{\gamma}(\mathbf{I} + \frac{\mathbf{G}}{\lambda})^{\lambda\tau} > 0$ for $\lambda > \lambda'$, because the left-hand side is a strictly monotone decreasing function of $\lambda$. Note that $\lambda'$ is explicitly computable from (4).

Next we investigate the sign of the rest of the elements of vector $\boldsymbol{\gamma}\mathbf{W}$. We apply Lemma 8 again, this time for $\mathbf{H} = \frac{k\mathbf{G}}{\lambda}$ and $n = k$ to get

$$\left\|e^{\frac{k\mathbf{G}}{\lambda}} - \left(\mathbf{I} + \frac{\mathbf{G}}{\lambda}\right)^k\right\| \leq e^{\frac{kg}{\lambda}}\frac{(kg)^2}{2k\lambda^2} \leq e^{\tau g}\frac{\tau g^2}{2\lambda}$$

11

uniformly in $0 \le k \le \lambda\tau$.

Let $\epsilon_2 = \inf_{0 \le x \le \tau} f_X(x) = \inf_{0 \le x \le \tau} \boldsymbol{\gamma} e^{\mathbf{G}x}(-\mathbf{G})\mathbb{1}$. Since $f_X(0) > 0$ as a result of Step 3 in Section 3.3, $\epsilon_2$ is strictly positive, due to the positive density condition. Let $V_k$ be the $k$-th coordinate of $\boldsymbol{\gamma}\mathbf{W}$ associated with the Erlang tail in (3); that is,

$$V_k = \boldsymbol{\gamma}\left(\mathbf{I} + \frac{\mathbf{G}}{\lambda}\right)^k \frac{-\mathbf{G}\mathbb{1}}{\lambda}.$$

Then

$$\left|\lambda V_k - f_X(\tfrac{k}{\lambda})\right| = \left|\boldsymbol{\gamma}\left[e^{\frac{k\mathbf{G}}{\lambda}} - \left(\mathbf{I} + \frac{\mathbf{G}}{\lambda}\right)^k\right]\mathbf{G}\mathbb{1}\right| \le \|\boldsymbol{\gamma}\|\left\|e^{\frac{k\mathbf{G}}{\lambda}} - \left(\mathbf{I} + \frac{\mathbf{G}}{\lambda}\right)^k\right\|\|\mathbf{G}\|\|\mathbb{1}\| \le \|\boldsymbol{\gamma}\|e^{\tau g}\frac{\tau g^2}{2\lambda}g\|\mathbb{1}\|.$$

Define $\lambda''$ as the solution of

$$\|\boldsymbol{\gamma}\|e^{\tau g}\frac{\tau g^2}{2\lambda}g\|\mathbb{1}\| = \epsilon_2. \tag{5}$$

$\lambda''$ is also explicitly computable. (Note that $\|\mathbb{1}\| = 1$, see Appendix A). For all $\lambda > \lambda''$ we have $V_k > 0$ because $f_X(\tfrac{k}{\lambda}) \ge \epsilon_2$ and the difference between $\lambda V_k$ and $f_X(\tfrac{i}{\lambda})$ is less than $\epsilon_2$.

Putting these together, we get that for $\tau$ and $\lambda = \max(\lambda', \lambda'')$ both parts of the vector $\boldsymbol{\gamma}\mathbf{W}$, that is, $\boldsymbol{\gamma}\left(\mathbf{I} + \frac{\mathbf{G}}{\lambda}\right)^n$ and $\boldsymbol{\gamma}\left(\mathbf{I} + \frac{\mathbf{G}}{\lambda}\right)^k \frac{-\mathbf{G}\mathbb{1}}{\lambda}$ for $k = 0, 1, \ldots, n-1$, are positive where $n = \lceil \tau\lambda \rceil$ and the obtained representation is indeed Markovian.

## 3.6 Step 5: correction related to Step 2

If Step 2 was applied, $(\boldsymbol{\beta}, \mathbf{B})$ is actually a Markovian representation for $f_Y$; Lemma 5 ensures that

$$\boldsymbol{\beta}' = \{1, 0, 0, \ldots, 0\}$$

$$\mathbf{B}' = \begin{pmatrix} -\mu & \mu & & \\ & \ddots & \ddots & \\ & & -\mu & \mu\boldsymbol{\beta} \\ & & & \mathbf{B} \end{pmatrix}$$

is a Markovian representation for $f_X(x) = f_Y(x) * g(l, \mu, x)$.

# 4 Worked example

Let

$$\boldsymbol{\alpha} = \frac{102}{139}\begin{pmatrix} 1 & 1 & -\frac{1}{3} & -\frac{2}{3} & -\frac{5}{2} & \frac{12}{17} & \frac{14}{17} \end{pmatrix},$$

$$\mathbf{A} = \begin{pmatrix} -1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 4 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -5 & 3 \\ 0 & 0 & 0 & 0 & 0 & -3 & -5 \end{pmatrix},$$

then

$$f(x) = -\boldsymbol{\alpha}\mathbf{A}e^{\mathbf{A}x}\mathbb{1} = \frac{102}{139}\left(xe^{-x} + e^{-x} + e^{-3x} - 10e^{-4x} + e^{-5x}\left(8\cos(3x) + 4\sin(3x)\right)\right).$$

The eigenvalues of $A$ are $-1$ (with multiplicity 2), $-3, -4, -5 + 3i, -5 - 3i$ and $1$. The eigenvalue 1 is redundant: the corresponding right-eigenvector is orthogonal to $\boldsymbol{\alpha}$, thus it does not appear in the pdf. It is eliminated during Step 1.

After Step 1, a minimal representation is obtained:

$$\boldsymbol{\alpha}_1 = \frac{102}{139} \begin{pmatrix} 1 & 1 & \frac{1}{3} & -\frac{5}{2} & \frac{13+\mathrm{i}}{17} & \frac{13-\mathrm{i}}{17} \end{pmatrix},$$

$$\mathbf{A}_1 = \begin{pmatrix} -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -3 & 0 & 0 & 0 \\ 0 & 0 & 0 & -4 & 0 & 0 \\ 0 & 0 & 0 & 0 & -5+3\mathrm{i} & 0 \\ 0 & 0 & 0 & 0 & 0 & -5-3\mathrm{i} \end{pmatrix}.$$

Since $f(0) = 0$, Step 2 needs to be applied.

$$f(0) = 0 \qquad f'(0) = 7 > 0,$$

so the value of $k$ in Lemma 4 is $k = 1$. Setting $\mu = 10$, the transformed pdf after Step 2 (borrowing the notation $f_Y$ from Lemma 4) is

$$f_Y(x) = \frac{102}{139} \left( \frac{9}{10} x e^{-x} + e^{-x} + \frac{7}{10} e^{-3x} - 6e^{-4x} + \frac{13+\mathrm{i}}{5} e^{(-5+3\mathrm{i})x} + \frac{13-\mathrm{i}}{5} e^{(-5-3\mathrm{i})x} \right)$$

and the corresponding representation for $f_Y$ is

$$\boldsymbol{\alpha}_2 = \frac{102}{139} \begin{pmatrix} \frac{9}{10} & 1 & \frac{2}{5} & -\frac{7}{15} & \frac{23}{340} & \frac{7}{340} \end{pmatrix},$$

$$\mathbf{A}_2 = \begin{pmatrix} -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -3 & 0 & 0 & 0 \\ 0 & 0 & 0 & -4 & 0 & 0 \\ 0 & 0 & 0 & 0 & -5+3\mathrm{i} & 0 \\ 0 & 0 & 0 & 0 & 0 & -5-3\mathrm{i} \end{pmatrix}.$$

From now on, we work with this representation. In Step 3, the eigenvalue pair $5 \pm 3\mathrm{i}$ is represented by a feedback-Erlang block. The order of this pair is $b = 4$, and the corresponding FE-block is

$$\begin{pmatrix} -5 & 5 & 0 & 0 \\ 0 & -5 & 5 & 0 \\ 0 & 0 & -5 & 5 \\ \frac{81}{125} & 0 & 0 & -5 \end{pmatrix}.$$

Step 3 results in the representation

$$\boldsymbol{\gamma} = \frac{102}{139} \begin{pmatrix} \frac{315}{2176} & \frac{10733}{21760} & \frac{6641}{32640} & \frac{8399}{21760} & \frac{147}{680} & -\frac{67}{272} & -\frac{45}{1088} & \frac{225}{1088} \end{pmatrix},$$

$$\mathbf{G} = \begin{pmatrix} -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -3 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -4 & 4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -5 & 5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -5 & 5 \\ 0 & 0 & 0 & 0 & \frac{81}{125} & 0 & 0 & -5 \end{pmatrix}.$$

Since $\boldsymbol{\gamma}$ still contains negative elements, Step 4 needs to be applied.
Following the algorithm in the proof of Theorem 7, we obtain the following values:

- $\tau = 0.5$ (from $\boldsymbol{\gamma} e^{\mathbf{G}\tau} > 0$),

- $g = \|G\| = 10$,

13

- $\|\boldsymbol{\gamma}\| < 1.5$,

- $\epsilon_1 > 0.05$ (for $\tau = 0.5$),

- $\lambda' = 112000$ from (4),

- $\epsilon > 0.069$, and thus $\lambda'' = 806600$ from (5).

This means that applying Step 4 with $\lambda = 806600$ and $n = \tau\lambda = 403300$ we obtain a Markovian representation for $f_Y$ in the form of (2).

Finally, Step 5 applies, so by Lemma 5 with $\mu = 10$ and $k = 1$, we obtain a Markovian representation for the original ME($\boldsymbol{\alpha}, \mathbf{A}$). The representation is of order 403309. Note that the order of this representation is *very* far from minimal, but we do not pursue a minimal value.

## 5    Conclusion

We have proposed a constructive proof for O'Cinneide's characterization theorem [7] along with an algorithm that always succeeds in finding a Markovian representation. The algorithm and the proof are divided into a few distinct steps, connecting some of the modern results in the field as well as introducing some original ideas using elementary function theory and matrix analysis.

## Acknowledgement

## References

[1] Peter Buchholz and Miklós Telek. On minimal representation of rational arrival processes. *Annals of Operations Research*, 202(1):35–58, 2013.

[2] C. Commault and S. Mocanu. Phase-type distributions and representations: some open problems for system theory. *Int. J. Control*, 76(6):566–580, 2003.

[3] R. S. Maier. The algebraic construction of phase-type distributions. *Commun. Stat., Stochastic Models*, 7(4):573 – 602, 1991.

[4] C. D. Meyer. *Matrix Analysis and Applied Linear Algebra*. SIAM, 2004.

[5] S. Mocanu and C. Commault. Sparse representations of phase-type distributions. *Commun. Stat., Stochastic Models*, 15(4):759 – 778, 1999.

[6] M.F. Neuts. *Matrix Geometric Solutions in Stochastic Models*. Johns Hopkins University Press, Baltimore, 1981.

[7] Colm Art OCinneide. Characterization of phase-type distributions. *Communications in Statistics. Stochastic Models*, 6(1):1–57, 1990.

[8] M. Soittola. Positive rational sequences. *Theoretical Computer Science*, 2:317–322, 1976.

[9] Kotaro Suzumura. Perron-frobenius theorem on non-negative square matrices: An elementary proof. *Hitotsubashi Journal of Economics*, 24(2):137–141, 1983.

[10] M. Telek and G. Horváth. A minimal representation of Markov arrival processes and a moments matching method. *Performance Evaluation*, 64(9-12):1153–1168, 2007.

# A Vector and matrix norms

We need some auxiliary facts about vector and matrix norms. First we define vector norms. Let $\boldsymbol{v}$ be a vector of size $n$.

**Definition 7.** *The 1-norm and $\infty$-norm of $\boldsymbol{v}$ are*

$$\|\boldsymbol{v}\|_1 = \sum_{i=1}^{n} |v_i|, \qquad \|\boldsymbol{v}\|_\infty = \max_{1 \le i \le n} |v_i|.$$

**Lemma 10.**     *a) $\|.\|_1$ and $\|.\|_\infty$ are equivalent, i.e,*

$$\|\boldsymbol{v}\|_\infty \le \|\boldsymbol{v}\|_1 \le n \, \|\boldsymbol{v}\|_\infty.$$

*b) If $\boldsymbol{v}$ is a row vector and $\boldsymbol{w}$ a column vector, then*

$$|\boldsymbol{vw}| \le \|\boldsymbol{v}\|_1 \, \|\boldsymbol{w}\|_\infty, \qquad |\boldsymbol{vw}| \le \|\boldsymbol{v}\|_\infty \, \|\boldsymbol{w}\|_1.$$

The fact that they are equivalent means that they define the same topology, so convergence to 0 is equivalent in either norm. For convenience, we will stick to using $\|.\|_1$ for row vectors and $\|.\|_\infty$ for column vectors.

We also need a matrix norm.

**Definition 8.** *The $\infty$-norm of $\mathbf{A}$ is*

$$\|\mathbf{A}\|_\infty = \max_{1 \le i \le n} \sum_{j=1}^{n} |A_{ij}|$$

This is a submultiplicative norm:

$$\|\mathbf{AB}\|_\infty \le \|\mathbf{A}\|_\infty \, \|\mathbf{B}\|_\infty.$$

Actually, the above matrix norm is the induced matrix norm of the vector norm $\|.\|_\infty$ when multiplying a column vector with a matrix from the left, and the induced matrix norm of the vector norm $\|.\|_1$ when multiplying a row vector with a matrix from the right. This means it works nicely with the previous vector norms.

**Lemma 11.** *Let $\boldsymbol{v}$ be a row vector and $\boldsymbol{w}$ a column vector of size $n$ and $\mathbf{A}$ be an $n \times n$ matrix. Then*

$$\|\boldsymbol{v}\mathbf{A}\|_1 \le \|v\|_1 \, \|\mathbf{A}\|_\infty, \quad \|\mathbf{A}\boldsymbol{w}\|_\infty \le \|\mathbf{A}\|_\infty \, \|\boldsymbol{w}\|_\infty, \quad |\boldsymbol{v}\mathbf{A}\boldsymbol{w}| \le \|\boldsymbol{v}\|_1 \, \|\mathbf{A}\|_\infty \, \|\boldsymbol{w}\|_\infty.$$

# B Proofs for the necessary direction

**Definition 9.** *The Markovian ($\boldsymbol{\alpha}$, $\mathbf{A}$) representation of PH($\boldsymbol{\alpha}$, $\mathbf{A}$) is* redundant *if it contains at least one state which cannot be visited by the Markov chain with initial distribution $\boldsymbol{\alpha}$ and generator $\mathbf{A}$. Otherwise ($\boldsymbol{\alpha}$, $\mathbf{A}$) is* non-redundant.

If the representation ($\boldsymbol{\alpha}$, $\mathbf{A}$) is redundant then it is easy to identify and eliminate the redundant states. Consider the vector $-\boldsymbol{\alpha}\mathbf{A}^{-1}$. The stochastic interpretation of its $i$th coordinate is the mean time spent in state $i$ before absorption. If the $i$th element of vector $-\boldsymbol{\alpha}\mathbf{A}^{-1}$ is zero then state $i$ is redundant and the associated elements can be deleted from vector $\boldsymbol{\alpha}$ and matrix $\mathbf{A}$ without modifying the distribution of time till absorption.

**Lemma 12.** *If $X$ is PH($\boldsymbol{\alpha}$,$\mathbf{A}$) distributed and non-redundant, then the positive density condition holds, that is,*

$$f_X(x) > 0 \qquad \forall x > 0.$$

*Proof.* If $X$ is PH($\boldsymbol{\alpha}$,$\mathbf{A}$) distributed and non-redundant then there is a path from every state with positive initial probability to the absorbing state and every state belongs to one of those paths. Consequently, the Markov chain is in state $j$ at time $x$ with positive probability, for any time $x > 0$ and for any state $j$. Let state $i$ be a transient state from where the absorption rate $g_i$ is positive. Then

$$f_X(x) = \boldsymbol{\alpha} e^{\mathbf{A}x}(-\mathbf{A})\mathbb{1} = \sum_{j=1}^n \Pr(Z(x) = j)g_j \geq Pr(Z(x) = i)g_i > 0,$$

where $Z(x)$ denotes the underlying Markov chain. $\qquad\square$

**Lemma 13.** *If $X$ is PH($\boldsymbol{\alpha}$,$\mathbf{A}$) distributed and non-redundant, then the dominant eigenvalue condition holds.*

Before proving Lemma 13, we elaborate on Definition 2. Let $ME(\boldsymbol{\gamma}, \mathbf{G})$ be a minimal representation for $X$. Consider its pdf using the Jordan-decomposition of $\mathbf{G}$ ($\mathbf{G} = \mathbf{P}\mathbf{J}\mathbf{P}^{-1}$)

$$f_X(x) = -\boldsymbol{\gamma}\mathbf{P}\mathbf{J}e^{\mathbf{J}x}\mathbf{P}^{-1}\mathbb{1} = \sum_{i=1}^l -\boldsymbol{\gamma}\mathbf{P}_i\mathbf{J}_i e^{\mathbf{J}_i x}\mathbf{P}_i'\mathbb{1},$$

where $\mathbf{J}_i$ denotes the Jordan-block corresponding to the eigenvalue $-\lambda_i$ and $\mathbf{P}_i$ denotes the submatrix of $\mathbf{P}$ containing only the columns corresponding to $\mathbf{J}_i$. $\mathbf{P}_i'$ denotes the submatrix of $\mathbf{P}^{-1}$ that contains only the rows corresponding to $\mathbf{J}_i$ (thus $\mathbf{P}_i$ is of size $n \times n_i$, where $n_i$ is the multiplicity of $-\lambda_i$ and $n$ is the size of $\mathbf{G}$, and $\mathbf{P}_i'$ is of size $n_i \times n$). In $\mathbf{P}_i$, the first column of each block is the (unique, up to a constant factor) right eigenvector $\boldsymbol{v}_i$ corresponding to that eigenvalue and the other columns are generalized eigenvectors. Similarly in $\mathbf{P}_i'$, the last row of each block is the (unique, up to a constant factor) left eigenvector $\boldsymbol{u}_i$ corresponding to that eigenvalue and the rest of the rows are generalized eigenvectors. If $i \neq j$, then $\mathbf{P}_i'\mathbf{P}_j = \mathbf{0}$.

The dominant term of $e^{\mathbf{J}_i x}$ is equal to $\frac{x^{n_i-1}e^{-\lambda_i x}}{(n_i-1)!}$ (where $n_i$ denotes the size of $\mathbf{J}_i$), and it is situated in the upper right corner. Within $-\boldsymbol{\gamma}\mathbf{P}_i\mathbf{J}_i e^{\mathbf{J}_i x}\mathbf{P}_i'\mathbb{1}$ this dominant term is obtained exactly when taking

$$-\boldsymbol{\gamma}\boldsymbol{v}_i\mathbf{J}_i e^{\mathbf{J}_i x}\boldsymbol{u}_i\mathbb{1} = (\boldsymbol{\gamma}\boldsymbol{v}_i)\lambda_i \frac{x^{n_i-1}e^{-\lambda_i x}}{(n_i-1)!}(\boldsymbol{u}_i\mathbb{1}).$$

If any of the coefficients $(\boldsymbol{\gamma}\boldsymbol{v}_i)$ and $(\boldsymbol{u}_i\mathbb{1})$ is 0, this term would vanish. Properties P3 and P4 ensure that this is not the case, in other words, all eigenvalues contribute to the pdf with maximal multiplicity (that is, Property P2).

This allows us to prove the DEC for any (possibly non-minimal) Markovian representation ($\boldsymbol{\alpha}, \mathbf{A}$) by proving that there exists a real eigenvalue of $\mathbf{A}$ that is strictly greater than the real part of all other eigenvalues AND this eigenvalue contributes to the pdf with maximal multiplicity.

The proof of Lemma 13 is based essentially on the Perron–Frobenius lemma. We begin by citing the Perron–Frobenius lemma along with a necessary definition, see for example [4].

**Definition 10.** *An $n \times n$ matrix $\mathbf{M}$ is* reducible *if there exists a nontrivial partition $I \cup J$ of $\{1, 2, \ldots, n\}$ such that*

$$\mathbf{M}_{ij} = 0 \qquad \forall i \in I, j \in J.$$

*Otherwise, $\mathbf{M}$ is* irreducible.

In case $\mathbf{M}$ is the transient generator of a PH distribution, then irreducibility means that each state can be reached from any other state before absorption, in this case we say that $\mathbf{M}$ has a single communicating class. If the Markov chain defined by $\mathbf{M}$ has multiple communicating classes, they correspond to a partition of the states as in the above definition.

**Theorem 14** (Perron–Frobenius). *If the irreducible matrix $\mathbf{M}$ has nonnegative elements, then there exists a positive eigenvalue $\nu_1$ of $\mathbf{M}$ such that*

- $\nu_1$ *has multiplicity 1,*

- $\nu_1 \geq |\nu_i| \, \forall i$ *where $v_i$ denote the eigenvalues of $\mathbf{M}$, and*

- *the corresponding right-eigenvector $\boldsymbol{v}_1$ is strictly positive (note that $\boldsymbol{v}_1$ is unique up to a constant factor; it can be chosen such that $\boldsymbol{v}_1$ is strictly positive).*

See Theorem 3 in [9] for a short, self-contained proof or Chapter 8 in [4] for a more detailed discussion. Note that the same conclusion holds for the left-eigenvector $\boldsymbol{u}_1$ as well. Note that the fact that $\nu_1$ is positive with multiplicity 1 and $\nu_1 \geq |\nu_i|$ mean that $\Re(\nu_i) < \nu_1$ for $i \neq 1$.

**Proof of Lemma 13.**

In case $\mathbf{A}$ has a single communicating class we apply Theorem 14 to the matrix $\mathbf{M} = \mathbf{A} + \omega\mathbf{I}$, where $\omega = \max_i |a_{ii}|$. Given that the matrix $\mathbf{A}$ is Markovian, $\mathbf{M}$ is nonnegative with the same eigenvectors and the eigenvalues shifted by $\omega$. The dominant eigenvalue $\nu_1$ of $\mathbf{M}$ corresponds to the dominant eigenvalue $-\lambda_1$ of $\mathbf{A}$, that is $\nu_1 = -\lambda_1 + \omega$ and the same relation holds for the other eigenvectors. Clearly for $i \neq 1$

$$\Re(\nu_i) < \nu_1 \quad \Longrightarrow \quad \Re(-\lambda_i) < -\lambda_1.$$

If $\mathbf{A}$ has a single communicating class then Theorem 14 guarantees that the multiplicity of $-\lambda_1$ is 1; this means that the unique dominant term in the pdf is $(\boldsymbol{\alpha}\boldsymbol{v}_1)\lambda_1 e^{-\lambda_1 x}(\boldsymbol{u}_1\mathbb{1})$. Strict positivity of $\boldsymbol{v}_1$ and $\boldsymbol{u}_1$ ensure $\boldsymbol{\alpha}\boldsymbol{v}_1 > 0$ and $\boldsymbol{u}_1\mathbb{1} > 0$, so indeed $\lambda_1$ contributes to the pdf with multiplicity 1.

If $\mathbf{A}$ has several communicating classes, the states can be renumbered such that $\mathbf{A}$ is an upper block triangular matrix, where each diagonal block corresponds to a communicating class and the blocks above the diagonal correspond to transitions between classes. The diagonal blocks are denoted by $\mathbf{B}_1, \ldots, \mathbf{B}_k$. The eigenvalues of $\mathbf{A}$ are the union of the eigenvalues associated with these diagonal blocks. Each $\mathbf{B}_i$ is itself the generator of a transient Markov chain, and, since $\mathbf{B}_i$ is also irreducible, Theorem 14 can be applied to each of them. It follows that each of these blocks (communicating classes) has its own dominant eigenvalue such that within that class, the real parts of all other eigenvalues are strictly smaller. It follows directly that the largest eigenvalue of $\mathbf{A}$ (denoted by $-\lambda_1$) is real and has $-\lambda_1 > \Re(-\lambda_i)$ for all $\lambda_i \neq \lambda_1$.

However, as opposed to the single class case, the multiplicity of $-\lambda_1$ may be higher than 1. Also, there may be several eigenvectors corresponding to $-\lambda_1$. This means that in order to calculate the contribution of $-\lambda_1$ to the pdf, we need to be slightly more meticulous. The proof is essentially a transformation of the matrix $\mathbf{A}$ to a form that is similar to the Jordan form (but not the same), while preserving some nonnegativity of $\mathbf{A}$ and $\boldsymbol{\alpha}$ (where it is important). We also present a numerical example at the end of this section to demonstrate the steps of the proof.

Let $\mathbf{Q}_i\mathbf{J}_i\mathbf{Q}_i^{-1} = \mathbf{B}_i$ be the Jordan decomposition of $\mathbf{B}_i$. We assume that the first block of $\mathbf{J}_i$ is the single dominant eigenvalue of $\mathbf{B}_i$; Theorem 14 thus guarantees that the first column of $\mathbf{Q}_i$, which is the corresponding right eigenvector, is strictly positive, and the first row of $\mathbf{Q}_i^{-1}$, which is the corresponding left eigenvector, is also strictly positive. Create the transformation matrix

$$\mathbf{Q} = \left[ \begin{array}{ccccc} \mathbf{Q}_1 & 0 & 0 & \ldots & 0 \\ 0 & \mathbf{Q}_2 & 0 & \ldots & 0 \\ \vdots & & & & \vdots \\ 0 & & \ldots & 0 & \mathbf{Q}_k \end{array} \right].$$

Then $\mathbf{Q}^{-1}\mathbf{A}\mathbf{Q}$ is an upper triangular matrix that contains the eigenvalues of $\mathbf{A}$ in its diagonal. Applying this transformation to the pdf, we get

$$f_X(x) = -\boldsymbol{\alpha}\mathbf{A}e^{\mathbf{A}x}\mathbb{1} = -(\boldsymbol{\alpha}\mathbf{Q})(\mathbf{Q}^{-1}\mathbf{A}\mathbf{Q})e^{(\mathbf{Q}^{-1}\mathbf{A}\mathbf{Q})x}(\mathbf{Q}^{-1}\mathbb{1}).$$

Take all rows and columns of $\mathbf{Q}^{-1}\mathbf{A}\mathbf{Q}$ that have $-\lambda_1$ in the diagonal. Denote this submatrix by $\mathbf{B}$. The submatrix $\mathbf{B}$ is responsible for the whole contribution of $-\lambda_1$. $\mathbf{B}$ can be calculated as

$$\mathbf{B} = \mathbf{R}\mathbf{Q}^{-1}\mathbf{A}\mathbf{Q}\mathbf{R}^T$$

where $\mathbf{R}$ is a $n_1 \times n$ binary matrix (whose elements are either 0 or 1) where $n_1$ is the multiplicity of the dominant eigenvalue in $\mathbf{A}$ and $n$ is the size of $\mathbf{A}$; row $i$ in $\mathbf{R}$ is equal to the unit vector $\boldsymbol{e}_j$ if the $i$-th instance of $-\lambda_1$ in the diagonal of $\mathbf{Q}^{-1}\mathbf{A}\mathbf{Q}$ is at coordinate $j, j$. $(\boldsymbol{\alpha}\mathbf{Q})$ is strictly positive on the coordinates corresponding to $\mathbf{B}$ since the dominant eigenvector of $\mathbf{Q}_i$ are strictly positive and the block of $\boldsymbol{\alpha}$ associated with $\mathbf{Q}_i$ is nonnegative and different from 0 (if it was 0 then $PH(\boldsymbol{\alpha},\mathbf{A})$ would be redundant). Similarly, $(\mathbf{Q}^{-1}\mathbb{1})$ is strictly positive on the coordinates corresponding to $\mathbf{B}$.

Finally, we argue that we can identify the dominant term in $e^{\mathbf{B}x}$ and see that it has a positive coefficient. This is done directly instead of transforming $\mathbf{B}$ to Jordan form. To this end, note that the offdiagonal elements of $\mathbf{B}$ are nonnegative since $\mathbf{A}$ originally contained nonnegative elements above the diagonal, which were then multiplied by the strictly positive dominant left and right eigenvectors of each block $\mathbf{B}_i$.

The matrix $\lambda_1\mathbf{I} + \mathbf{B}$ is strictly upper triangular, thus nilpotent; this implies that the series expansion

$$e^{(\lambda_1\mathbf{I}+\mathbf{B})x} = \sum_{k=0}^{\infty} \frac{((\lambda_1\mathbf{I} + \mathbf{B})x)^k}{k!}$$

is actually a finite sum, and $e^{(\lambda_1\mathbf{I}+\mathbf{B})x}$ is a polynomial of $x$. The dominant term in $e^{\mathbf{B}x}$ is equal to the last nonzero term of this polynomial, multiplied by $e^{-\lambda_1 x}$. The coefficient of this term is necessarily positive since $(\lambda_1\mathbf{I} + \mathbf{B})$ and thus powers of $(\lambda_1\mathbf{I} + \mathbf{B})$ do not have negative elements.

Consequently, we have proved that $\lambda_1$ contributes to the pdf

$$f_X(x) = -\boldsymbol{\alpha}\mathbf{A}e^{\mathbf{A}x}\mathbb{1} = -(\boldsymbol{\alpha}\mathbf{A})(\mathbf{Q}^{-1}\mathbf{A}\mathbf{Q})e^{(\mathbf{Q}^{-1}\mathbf{A}\mathbf{Q})x}(\mathbf{Q}^{-1}\mathbb{1}).$$

with maximal multiplicity and with a positive coefficient, and the DEC holds.

**Example 1.** *Let*

$$\mathbf{A} = \begin{bmatrix} -4 & 1 & 1 & 0 & 0.2 & 0.4 & 0 & 0 & 0 & 0.4 \\ 1 & -2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & -3 & 0 & 0 & 0 & 0 & 0.2 & 0.4 & 0.2 \\ 0 & 0 & 0 & -4 & 3 & 0.2 & 0.2 & 0 & 0.4 & 0 \\ 0 & 0 & 0 & 1 & -2 & 0 & 0.2 & 0.2 & 0 & 0.2 \\ 0 & 0 & 0 & 0 & 0 & -2 & 1 & 0 & 1/5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -8 & 2 & 0.6 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 6 & -7 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix}.$$

$\mathbf{A}$ *has 5 communicating classes:* $\mathbf{B}_1$ *has size 3 and dominant eigenvalue* $-1$, $\mathbf{B}_2$, $\mathbf{B}_3$ *and* $\mathbf{B}_4$ *are of size 2 and their dominant eigenvalues are* $-1, -1$ *and* $-4$ *respectively;* $\mathbf{B}_5$ *is of size 1 with dominant eigenvalue* $-1$. *Thus* $\lambda_1 = 1$.

$$\mathbf{Q} = \begin{bmatrix} 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

*Notice that in* $\mathbf{Q}$*, the first column in each block is strictly positive. Even though it is not displayed in this example,* $\mathbf{Q}$ *(and* $\mathbf{Q}^{-1}\mathbf{A}\mathbf{Q}$*) may contain complex numbers, but only in rows and columns corresponding to non-dominant eigenvalues.*

$$\mathbf{Q}^{-1}\mathbf{A}\mathbf{Q} = \begin{bmatrix} -1 & 0 & 0 & 0.05 & 0.05 & 0.10 & -0.10 & 0.25 & 0.20 & 0.15 \\ 0 & -3 & 0 & 0.10 & 0.10 & 0.20 & -0.20 & 0.50 & 0.40 & 0.30 \\ 0 & 0 & -5 & -0.15 & 0.30 & -0.15 & -0.30 & 0.25 & 0.20 & -0.25 \\ 0 & 0 & 0 & -1 & 0 & 0.25 & 0.15 & 0.35 & 0 & 0.15 \\ 0 & 0 & 0 & 0 & -5 & -0.05 & 0.05 & -0.15 & -0.40 & 0.05 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0.20 & 0.30 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -3 & -0.20 & -0.30 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -4 & 0 & 9/35 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -11 & -6/35 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix}.$$

*The rows and columns that include the dominant eigenvalue are marked and so*

$$\mathbf{R} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{B} = \mathbf{R}\mathbf{Q}^{-1}\mathbf{A}\mathbf{Q}\mathbf{R}^T = \begin{bmatrix} -1 & 0.05 & 0.10 & 0.15 \\ 0 & -1 & 0.25 & 0.15 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}.$$

*The last nonzero power of the nilpotent matrix $\lambda_1 \mathbf{I} + \mathbf{B}$ is*

$$(\lambda_1 \mathbf{I} + \mathbf{B})^2 = \begin{bmatrix} 0 & 0 & 0.00125 & 0.0075 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

*whose nonzero elements are all positive.*

# C    Proofs for the sufficient direction

## C.1    Proof of Lemma 4.

The intuitive behavior of the convolution of the pdf of a non-negative r.v. $(Y)$ and the Erlang$(l, \mu)$ pdf is the following: assume $f_Y(0) > 0$; for large values of $\mu$, the Erlang pdf decays rapidly, so the function $f_Y$ is very close to $f_X$, except around 0, since convolution of a pdf $f_Y$ with an Erlang$(l, \mu)$ pdf increases the multiplicity of 0 by $l$. Lemma 4 utilizes this relation in the opposite direction. Hence if $f_X$ was positive everywhere except at 0 with multiplicity $l$, then $f_Y$ will be positive at 0 and its positivity for $\mathbb{R}^+$ comes from the small difference from $f_X$. (Actually, the tail and the main body of $f_Y(x)$ will be examined separately for technical reasons.)

$f_Y$ can be calculated in the Laplace-transform domain as follows. The Laplace-transform of the Erlang$(l, \mu)$ pdf is

$$f^*_{k,\mu}(s) = \left( \frac{\mu}{s + \mu} \right)^l.$$

Denote by $f^*_X(s)$ and $f^*_Y(s)$ the Laplace-transform of $f_X$ and $f_Y$, respectively. Then from $f_X(x) = f_Y(x) * f_{l,\mu}(x)$ we have $f^*_X(s) = f^*_Y(s) \cdot \left( \frac{\mu}{s+\mu} \right)^l$, and so

$$f^*_Y(s) = f^*_X(s) \left( \frac{s + \mu}{\mu} \right)^l = f^*_X(s) \left( 1 + \frac{s}{\mu} \right)^l.$$

For $l = 1$, the inverse transform of $f^*_X(s) \left( 1 + \frac{s}{\mu} \right)$ gives

$$f_Y(x) = f_X(x) + \frac{1}{\mu} \left( f'_X(x) + f_X(0) \right) = f_X(x) + \frac{1}{\mu} f'_X(x).$$

For $l > 1$, induction (or the binomial theorem) gives

$$f_Y(x) = \sum_{i=0}^{l} \binom{l}{i} \frac{1}{\mu^i} f_X^{(i)}(x) = -\gamma \sum_{i=0}^{l} \binom{l}{i} \left( \frac{\mathbf{G}}{\mu} \right)^i \mathbf{G} e^{\mathbf{G}x} \mathbb{1}.$$

The fact that $f_Y(x)$ is a matrix-exponential pdf is straightforward from the above formula. Also, it has a representation of the form $\text{ME}(\boldsymbol{\gamma}', \mathbf{G})$ where $\boldsymbol{\gamma}' = -\gamma \sum_{i=0}^{l} \binom{l}{i} \left( \frac{\mathbf{G}}{\mu} \right)^i$.

We fix a value $\delta > 0$ (independent from $\mu$) such that

$$f_X^{(l)}(x) > 0, \quad x \in (0, \delta].$$

This is possible since $f_X^{(l)}(0) > 0$ and $f_X^{(l)}$ is continuous. This in turn implies by integration that

$$f_X^{(i)}(x) \geq 0, \quad x \in (0, \delta].$$

for every $i = l, l-1, \ldots, 1, 0$ and thus

$$f_Y(x) = \sum_{i=0}^{l} \binom{l}{i} \frac{1}{\mu^i} f_X^{(i)}(x) > 0 \quad x \in (0, \delta).$$

19

This holds for any value of $\mu$.

We examine the tail of $f_Y$ next. Recall that as $x \to \infty$, $f_X(x)$ decays as $c_{\lambda_1,n_1} x^{n_1-1} e^{-\lambda_1 x}$ where $c_{\lambda_1,n_1} > 0$.

$$f_Y(x) - f_X(x) = \sum_{i=0}^{l} \binom{l}{i} \frac{1}{\mu^i} f_X^{(i)}(x) - f_X(x) = -\gamma \sum_{i=1}^{l} \binom{l}{i} \left(\frac{\mathbf{G}}{\mu}\right)^i \mathbf{G} e^{\mathbf{G}x} \mathbb{1}. \tag{6}$$

Since $e^{\mathbf{G}x}$ decays with rate $x^{n_1-1} e^{-\lambda_1 x}$,

$$-\gamma \binom{l}{i} \mathbf{G}^i \mathbf{G} e^{\mathbf{G}x} \mathbb{1} \sim c_i x^{n_1-1} e^{-\lambda_1 x}$$

for each $i = 1, \ldots, l$ for some constants $c_i$.

Select $K_1$ such that

$$\left| \frac{-\gamma \binom{l}{i} \mathbf{G}^i \mathbf{G} e^{\mathbf{G}x} \mathbb{1}}{x^{n_1-1} e^{-\lambda_1 x}} \right| \leq 2|c_i| \quad \forall x > K_1$$

for $i = 1, \ldots, k$, Then

$$|f_Y(x) - f_X(x)| \leq \sum_{i=1}^{l} \frac{2|c_i|}{\mu^i} x^{n_1-1} e^{-\lambda_1 x}.$$

Note that $K_1$ is also independent from $\mu$.

The constant $\sum_{i=1}^{l} \frac{2|c_i|}{\mu^i}$ is decreasing in $\mu$. Select $\mu_0$ such that

$$\sum_{i=1}^{l} \frac{2|c_i|}{\mu^i} \leq \frac{1}{2} c_{\lambda_1,n_1} \quad \forall \mu > \mu_0.$$

Select $K_2$ such that

$$f_X(x) \geq \frac{1}{2} c_{\lambda_1,n_1} x^{n_1-1} e^{-\lambda_1 x} \quad \forall x > K_2.$$

Set $K = \max(K_1, K_2)$. At this point, $\delta$ and $K$ are fixed (independently of $\mu$), and for any $\mu > \mu_0$ it holds that

$$f_Y(x) \geq f_X(x) - |f_Y(x) - f_X(x)| \geq \frac{1}{2} c_{\lambda_1,n_1} x^{n_1-1} e^{-\lambda_1 x} - \frac{1}{2} c_{\lambda_1,n_1} x^{n_1-1} e^{-\lambda_1 x} = 0 \quad \forall x > K.$$

We now have positivity of $f_Y$ at $[0, \delta]$ and $[K, \infty]$. For $[\delta, K]$, we use the formula (6) again, and note that

$$\sup_{x \in [\delta, K]} \left| \gamma \sum_{i=1}^{l} \binom{l}{i} \left(\frac{\mathbf{G}}{\mu}\right)^i \mathbf{G} e^{\mathbf{G}x} \mathbb{1} \right| \leq \sum_{i=1}^{l} \left(\frac{1}{\mu}\right)^i \sup_{x \in [\delta, K]} \left| \gamma \binom{l}{i} \mathbf{G}^i \mathbf{G} e^{\mathbf{G}x} \mathbb{1} \right|,$$

where $\sup_{x \in [\delta, K]} \left| \gamma \binom{l}{i} \mathbf{G}^i \mathbf{G} e^{\mathbf{G}x} \mathbb{1} \right|$ is finite for each $i = 1, \ldots, l$, while $\frac{1}{\mu^i} \to 0$, so there exists a $\mu_1$ such that for any $\mu > \mu_1$,

$$|f_Y(x) - f_X(x)| \leq \inf_{x \in [\delta, K]} f_X(x),$$

which is positive due to the positive density condition (specifically that $f_X$ is strictly positive on a finite interval not containing 0).

Selecting any $\mu > \max(\mu_1, \mu_2)$ finishes the lemma.

## C.2   Proof of Lemma 8.

We will prove part $i)$ first.

   We will begin by showing that the supremum is obtained for $z = r$.

   Series expansion gives

$$e^z - \left(1 + \frac{z}{n}\right)^n = \sum_{k=0}^{\infty} \frac{z^k}{k!} B(n, k),$$

where

$$B(n, k) = \begin{cases} 1 - \frac{n(n-1)\ldots(n-k+1)}{n^k} & \text{if } k \leq n \\ 1 & \text{if } k > n \end{cases}$$

Note the following properties of $B(n, k)$:

$$0 \leq B(n, k) \leq 1 \ \forall n, k; \qquad \lim_{n \to \infty} B(n, k) = 0 \ \forall k.$$

For every $z$ with $|z| \leq r$, we have

$$\left| e^z - \left(1 + \frac{z}{n}\right)^n \right| = \left| \sum_{k=0}^{\infty} \frac{z^k}{k!} B(n, k) \right| \leq \sum_{k=0}^{\infty} \frac{|z|^k}{k!} B(n, k) \leq \sum_{k=0}^{\infty} \frac{r^k}{k!} B(n, k) = \left| e^r - \left(1 + \frac{r}{n}\right)^n \right|.$$

Notice that the series expansion ensures $e^r - \left(1 + \frac{r}{n}\right)^n > 0$, so we only need an upper bound on $e^r - \left(1 + \frac{r}{n}\right)^n$. Using the straightforward inequalities

$$\ln(1 + x) \geq x - \frac{x^2}{2} \ (x \geq 0) \quad \text{and} \quad e^x \geq 1 + x \ (x \in \mathbb{R})$$

we get that

$$e^r - \left(1 + \frac{r}{n}\right)^n = e^r - e^{n \ln(1 + r/n)} \leq e^r - e^{r - r^2/(2n)} = e^r \left(1 - e^{-r^2/(2n)}\right) \leq e^r \left(1 - \left(1 - \frac{r^2}{2n}\right)\right) = e^r \frac{r^2}{2n}.$$

We note that this estimate is asymptotically sharp as $n \to \infty$.

For part $ii)$, we use the series expansion again:

$$\left\| e^{\mathbf{H}} - \left(1 + \frac{\mathbf{H}}{n}\right)^n \right\| = \left\| \sum_{k=0}^{\infty} \frac{\mathbf{H}^k}{k!} B(n, k) \right\| \leq \sum_{k=0}^{\infty} \frac{\|\mathbf{H}\|^k}{k!} B(n, k) \leq$$

$$\leq \sum_{k=0}^{\infty} \frac{r^k}{k!} B(n, k) = e^r - \left(1 + \frac{r}{n}\right)^n \leq \frac{r^2 e^r}{2n},$$

where $r = \|\mathbf{H}\|$.

## C.3   Proof of Lemma 9.

According to the FE block composition of $\mathbf{G}$ it has the following block structure

$$\mathbf{G} = \left[ \begin{array}{c|c} \mathbf{G}_{11} & \mathbf{G}_{12} \\ \hline 0 & \mathbf{G}_{22} \end{array} \right], \tag{7}$$

where

$$\mathbf{G}_{11} = \begin{bmatrix} -\lambda_1 & \lambda_1 & 0 & \ldots & 0 \\ 0 & -\lambda_1 & \lambda_1 & \ldots & 0 \\ \vdots & & & & \\ 0 & & \ldots & 0 & -\lambda_1 \end{bmatrix}, \quad \mathbf{G}_{12} = \begin{bmatrix} 0 & 0 & \ldots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \ldots & 0 \\ \lambda_1 & 0 & \ldots & 0 \end{bmatrix},$$

and $\mathbf{G}_{22}$ contains the rest of the FE blocks. The size of $\mathbf{G}_{11}$ is denoted by $n_1$ (which is the multiplicity of the dominant eigenvalue $-\lambda_1$) and the size of $\mathbf{G}_{22}$ by $n_2$. Let

$$\mathbf{H} = \mathbf{G} + \lambda_1 \mathbf{I},$$

and accordingly $\mathbf{H}_{11} = \mathbf{G}_{11} + \lambda_1 \mathbf{I}, \mathbf{H}_{12} = \mathbf{G}_{12}$ and $\mathbf{H}_{22} = \mathbf{G}_{22} + \lambda_1 \mathbf{I}$, where $\mathbf{I}$ denotes the identity matrix of appropriate size. From $\mathbf{H} = \mathbf{G} + \lambda_1 \mathbf{I}$, it follows that

$$e^{\mathbf{G}x} = e^{-\lambda_1 x} e^{\mathbf{H}x},$$

and it is enough to investigate the dominant row of $e^{\mathbf{H}x}$. In the rest of the proof, $(.)_{11}, (.)_{12}, (.)_{22}$ denote the corresponding matrix blocks (*not* single elements). The eigenvalues of $\mathbf{H}_{22}$ have negative real parts. Their real parts are less than or equal to $\lambda_1 - \Re(\lambda_2)$, where $-\lambda_2$ is the eigenvalue with the second largest real part.

From the series expansion of $e^{\mathbf{H}x}$

$$e^{\mathbf{H}x} = \sum_{n=0}^{\infty} \frac{x^n}{n!} \mathbf{H}^n,$$

and the block triangular structure of $\mathbf{H}$ we have that the upper left block is

$$(e^{\mathbf{H}x})_{11} = \sum_{n=0}^{\infty} \frac{x^n}{n!} \mathbf{H}_{11}^n,$$

where $(\mathbf{H}_{11}x)^n$ can be calculated explicitly:

$$(\mathbf{H}_{11}x)^n = \begin{bmatrix} 0 & \cdots & 0 & (\lambda_1 x)^n & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & (\lambda_1 x)^n & \cdots & 0 \\ \vdots & & & & & & \\ 0 & & & \cdots & & & (\lambda_1 x)^n \\ 0 & & & & & & 0 \\ \vdots & & & & & & \\ 0 & & & \cdots & & & 0 \end{bmatrix},$$

with the nonzero elements being at positions $(1, n+1), (2, n+2), \ldots$. Specifically, $\mathbf{H}_{11}^n$ is 0 for $n \geq n_1$, so the sum $\sum_{n=0}^{\infty} \frac{x^n}{n!} \mathbf{H}_{11}^n$, is actually finite, and from the above form it is clear that $(e^{\mathbf{H}t})_{11}$ is upper diagonal, dominated by its first row, which of course also dominates $(e^{\mathbf{H}x})_{21} = 0$.

The rest of the proof is devoted to the elements of $(e^{\mathbf{H}x})_{12}$ and $(e^{\mathbf{H}x})_{22}$. For that, we need to examine $(e^{\mathbf{H}x})_{12}$.0

$$(e^{\mathbf{H}x})_{12} = \sum_{n=0}^{\infty} \frac{x^n}{n!} (\mathbf{H}^n)_{12}.$$

Here,

$$(\mathbf{H}^n)_{12} = \sum_{k=0}^{n-1} (\mathbf{H}_{11})^k \mathbf{H}_{12} (\mathbf{H}_{22})^{n-k-1}$$

since $\mathbf{H}$ is an upper block bi-diagonal matrix. Thus

$$(e^{\mathbf{H}x})_{12} = \sum_{n=1}^{\infty} \frac{x^n}{n!} \sum_{k=0}^{n-1} (\mathbf{H}_{11})^k \mathbf{H}_{12} (\mathbf{H}_{22})^{n-k-1}$$

$$= \sum_{k=0}^{\infty} (\mathbf{H}_{11})^k \mathbf{H}_{12} \sum_{n=k+1}^{\infty} \frac{x^n}{n!} (\mathbf{H}_{22})^{n-k-1}$$

$$= \sum_{k=0}^{n_1-1} (\mathbf{H}_{11})^k \mathbf{H}_{12} \sum_{n=k+1}^{\infty} \frac{x^n}{n!} (\mathbf{H}_{22})^{n-k-1}.$$

Again, the sum over $k$ is finite.

The inner sum can be calculated as

$$\sum_{n=k+1}^{\infty} \frac{1}{n!} x^{n-k-1} = x^{-k-1} \sum_{n=k+1}^{\infty} \frac{1}{n!} x^n = x^{-k-1} \left( e^x - \sum_{l=0}^{k} \frac{x^l}{l!} \right),$$

and accordingly,

$$\sum_{n=k+1}^{\infty} \frac{x^n}{n!} (\mathbf{H}_{22})^{n-k-1} = (\mathbf{H}_{22})^{-(k+1)} \left( e^{x\mathbf{H}_{22}} - \mathbf{I} - \mathbf{H}_{22}x - \cdots - \frac{(\mathbf{H}_{22}x)^k}{k!} \right).$$

Putting it all together, we obtain that

$$(e^{\mathbf{H}x})_{12} = \sum_{k=0}^{n_1-1} (\mathbf{H}_{11})^k \mathbf{H}_{12} (\mathbf{H}_{22})^{-(k+1)} \left( e^{x\mathbf{H}_{22}} - \mathbf{I} - \mathbf{H}_{22}x - \cdots - \frac{(\mathbf{H}_{22}x)^k}{k!} \right).$$

The form of $(\mathbf{H}_{11})^k \mathbf{H}_{12}$ guarantees that for each $k$

$$(\mathbf{H}_{11})^k \mathbf{H}_{12} (\mathbf{H}_{22})^{-(k+1)} \left( e^{x\mathbf{H}_{22}} - \mathbf{I} - \mathbf{H}_{22}x - \cdots - \frac{(\mathbf{H}_{22}x)^k}{k!} \right)$$

has a single nonzero row, with $k = n_1 - 1$ corresponding to the first row being nonzero, $k = n_1 - 2$ to the second etc. Within each row, the main term is

$$-(\mathbf{H}_{11})^k \mathbf{H}_{12} (\mathbf{H}_{22})^{-(k+1)} \frac{(\mathbf{H}_{22}x)^k}{k!} = -\frac{x^k}{k!} (\mathbf{H}_{11})^k \mathbf{H}_{12} (\mathbf{H}_{22})^{-1}.$$

Specifically, the main term in each element of the first row is of order $x^{n_1-1}$, and the order in the other rows is smaller within the block $(e^{\mathbf{H}x})_{12}$.

We need to calculate $\mathbf{H}_{22}^{-1}$. It can be calculated either via Cramer's (which allows for calculating the constants $C_j$ explicitly, but is left to the reader), or by using the following identity:

$$\mathbf{H}_{22}^{-1} = -\int_{\tau=0}^{\infty} e^{\mathbf{H}_{22}x} \mathrm{d}t = -\int_{\tau=0}^{\infty} e^{\lambda_1 x} \cdot e^{\mathbf{G}_{22}x} \mathrm{d}t.$$

The integral exists because all eigenvalues of $\mathbf{H}_{22}$ have negative real part. $e^{\lambda_1 x}$ is a positive function ("weight") and $e^{\mathbf{G}_{22}x}$ contains the transition probabilities of a CTMC, so all elements of $e^{\mathbf{G}_{22}x}$ are positive for all $t > 0$. Thus all elements of $\mathbf{H}_{22}^{-1}$ are negative, and the single nonzero row of $-(\mathbf{H}_{11})^k \mathbf{H}_{12} (\mathbf{H}_{22})^{-(k+1)} \frac{(\mathbf{H}_{22}x)^k}{k!}$ is strictly positive.

Finally, since the block $(\mathbf{H})_{22}$ has eigenvalues with negative real part, the elements of $(e^{\mathbf{H}x})_{22}$ decay exponentially, so they are of course dominated by the first row of $(e^{\mathbf{H}x})_{12}$.

One last remark: [5, page 771] discusses the same statement in a rather descriptive manner using communicating classes.